Extending the Limits of Direct High Angular Resolution Infrared Astronomical Imaging

Richard Michael Brockie

Doctor of Philosophy
The University of Edinburgh
1998

This thesis is my own composition except where indicated in the text.

October 29, 1998

Acknowledgements

The Royal Observatory, Edinburgh is a great place. Its wonderful location has made coming here a joy every day and must be at least partly responsible for the great atmosphere among those who work here. Much kudos to the 26^{th} Earl of Crawford who made this possible.

Of the grown–ups, I would like to especially thank Martyn Wells, my principal supervisor, for guidance, insightful discussions and enthusiastically answering every question I have had. Thanks also to Suzie Ramsey Howat and John Cooke, my other supervisors, Andy Taylor for answering a staggering number of questions about statistics and matrices and the computing staff who have kept the system running most of the time.

Further afield, I would like to thank George Aitken of Queen's University, Ontario for an exciting collaboration on the prediction stuff, the UKIRT staff, in particular Gillian Wright for being a great support astronomer both times I was there and Leslie Lamport for developing LaTeX which has been a delight to use whilst writing this thesis.

The less grown–up inhabitants of ROE have proved to be much fun, especially the many and varied inhabitants of R7, perhaps *the* choice piece of student real estate at the Observatory, whose gay banter and silly games have lightened the mood: Hugh, Henry, Mike, Bill, Richard, Matt and Ross. Thanks also to Alison and Nick for being great compatriots.

Beyond astronomy, I would like to thank John Kitchen for letting me sing in his choir for so long, and the various members of the Edinburgh University Cycling Club for fun and pain. I would also like to not thank the driver who caused me to go to hospital rather than go on my first observing run.

This section would not be complete without acknowledging the varied sources of funds I have had whilst I have been here. They are, PPARC (studentship & travel), NATO (travel grant CRG–951322 to collaborate with George Aitken), Fiona Plows (not leaving me in the gutter), The University of Edinburgh (demonstrating and tutoring), ROE/UK ATC (webpages and Gemini Acquisition & Guidance System), Benefits Agency (erroneously paid Jobseekers' Allowance) and the insurance company of the driver who put me in hospital (loss of earnings and personal injury).

Finally, I would like to thank Fiona for being Fiona.

Abstract

Observing in the infrared (IR) part of the electromagnetic spectrum is now an established tool of astronomy. It allows investigations of, among others, high redshift galaxies, star formation regions and very low mass stars close to the hydrogen burning limit as well as providing information complementary to that obtained in other regions of the spectrum.

The dimensions of infrared arrays have increased over the years from 62×58 in IRCAM1, the first infrared imager on the UK Infrared Telescope, to the 256^2 array in IRCAM3, the current camera, soon to be superceded by 1024^2 arrays in the next generation of instruments. In this thesis, I describe the first observing programme which uses infrared observations to measure trigonometric parallaxes — made possible through the introduction of larger IR arrays. In this programme, certain difficulties associated with infrared techniques are encountered and described with results presented for a previously measured star and a brown dwarf candidate.

A major benefit of observing in the infrared is that atmospheric distortion has less of an effect on the formation of images — seeing on a good site can be

$< 0.5''$ at $2\mu m$. The recent development of Adaptive Optics (AO) systems, which compensate for wavefront aberrations as observations are made, further reduce the effects of atmospheric distortion.

AO systems have a servo–loop in which a deformable mirror attempts to remove the distortion present in the measured wavefront. In this thesis, I describe a method of real time characterisation of the most recent behaviour of the atmosphere, as observed by an AO system. Rather than reacting to the last measured distortion, this knowledge can be used in the servo–loop to reduce mirror fitting errors by predicting the next mirror shape. I describe a series of simulations which prove the validity of this novel technique. Finally, with simulations of the AO system being built for the William Herschel Telescope, I show that the improvement in performance available through prediction allows use of an AO guide star about 0.25 magnitudes fainter when compared with the non–predictive case.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The human race is fixated with extremes. Whether it is the person who can cover a fixed distance in the quickest possible time, submerging to record depths in the oceans, or cooking the biggest pancake ever recorded, people strive to be exceptional. My aunt, Carol Brockie, and her friend Marie Warrell once held the world record of 110 hours for the longest darts marathon (Guinness, 1981).

The field of astronomy is not immune to these urges. Some people quest to discover the most distant quasar and others wish to perform the n–body simulation with the greatest number of particles ever fitted into a computer's memory. The locations of astronomical telescopes are extreme – high atop the peaks of dormant (not extinct!) volcanos, upon islands in the middle of the largest ocean on the globe so that as little of the atmosphere is above the telescope as possible and that what remains is as smooth as possible. In the last 20 years or so, this urge has been taken to its logical extreme with the launching of numerous telescopes and other astronomical satellites to remove the effects of the atmosphere altogether.

But we are not finished yet! The environment of Earth orbit is not good enough if we are to be able to detect what we think is out there, and recent proposals for the NGST (the proposed successor to the Hubble Space Telescope) have indicated that the telescope will be located well away from the Earth where the radiation environment is more benign.

This thesis is concerned with two such urges. First, the desire to unambiguously detect the smallest extra–solar system object ever detected, and second, the wish to image through the atmosphere as clearly as possible. Upon first inspection, these topics seem disparate, but allow me to convince you otherwise.

As light from a heavenly object passes through the atmosphere, it is distorted by turbulent air and as a result, spreads out as it passes through the atmosphere. When an image is formed of the star or galaxy, the area illuminated by the object on the photographic plate or semi–conductor detector sensitive to light is larger than if the atmosphere were not distorting the light. If the intensity per unit area drops below a certain limit, the signal caused by the light is masked by the noise in the detector. So, if the light is spread out too much, the object will not be detected.

The more massive a star is, the more light it emits, and conversely a star which is less massive emits less light. Less massive stars are also cool and consequently relatively strong emitters in the infrared. If our detector is serendipitously looking in the correct direction to a very faint star, but the atmosphere is smearing out the

light from the object, the amount of light falling in each element of the detector may not be enough to detect the object. However, if the smearing effect of the atmosphere can be removed in some fashion, there is more chance of detecting the faint object which may just turn out to be the most distant quasar ever discovered or a nearby brown dwarf below the hydrogen burning mass–limit.

At first inspection, a faint source could be a bright object a long way off, or a faint object close by. This degeneracy is broken through some estimation of the distance to the object. The fundamental method of measuring astronomical distances, and the first rung of the cosmological distance ladder is trigonometric parallax (Rowan-Robinson, 1985).

Chapter 2 of this thesis describes the first infrared observing programme to measure trigonometric parallaxes. Parallax observations in this wavelength region have only recently become possible through the development of infrared array detectors with enough pixels to provide both good resolution and a reasonably wide field of view.

Chapters 3 and 4 of this thesis are concerned with an investigation into reducing the effect of atmospheric smearing as much as possible. The last 20 or so years has seem the development and eventual use of a system called Adaptive Optics (AO) whilst performing astronomical observations. Such a system provides a mechanism to measure and correct *in real time* for the distortions introduced by the propagation of light through the atmosphere, allowing sharper images to be

formed. Adaptive Optics performs best when correcting infrared wavelengths.

Chapter 3 outlines the problems of atmospheric turbulence. The two flavours of adaptive optics systems are described, along with the general limitations of AO systems. Chapter 3 goes on to describe a novel method for characterising the most recent effects of the atmospheric turbulence. This information can be used to predict the likely behaviour of the atmosphere in the near future, and hence allow even sharper images to be formed than are possible without prediction.

Chapter 4 describes a series of computer simulations of an AO system, comparing the system on its own and with the additional predictive capability. Running the simulation as a model of the NAOMI AO system, under construction for the William Herschel Telescope on La Palma at various locations around the UK, shows that the guide star which provides the illumination to measure the atmospheric distortion can be a fifth fainter when using the predictor, compared with that used by the unaided AO system, for the same degree of atmospheric correction. This is good news: there are far more faint stars than bright ones, allowing more of the sky to be observed whilst using an AO system.

Chapter 5 summarises and concludes the thesis with some pointers towards future work.

Chapter 2

Infrared Astrometry

2.1 Introduction

Astrometry is the precise measurement of the position of astronomical objects. Through astrometric measurements, the parallax of a star may be determined. This method provides the fundamental basis for all distance measures used in astronomy by providing the bottom rung of the cosmological *distance ladder* (Rowan-Robinson, 1985).

The parallax of an astronomical object is the apparent motion against fixed background stars caused by the orbit of the Earth round the Sun. The distance of an object from Earth is inversely proportional to its parallax. Parallax is described in more detail in section 2.2. In addition to parallax, objects can exhibit *proper motion* with respect to reference stars due to the relative velocity between the Sun and the object.

Parallaxes have traditionally been measured using photographic plates (van Altena, 1983). The good resolution and large field of view afforded by photograph allows well–sampled images of not only the target star, but enough other stars with which to construct a coordinate system against which an estimate of the parallax may be made.

There are however two major drawbacks to photography as a technique. First, the detective quantum efficiency of the photographic emulsion is poor, in the region of only a few percent, *e.g.* $\sim 3\%$. This means that $\sim 97\%$ of the photons from the source are not detected by the emulsion, requiring long integrations for usable images to be obtained.

Second, the information contained in a photographic plate is not readily accessible. Before any analysis can be performed on the data contained within the images, the plates themselves must be measured. For example, van Altena (1971) describes a measuring process using a machine which required the operator to decide by eye when an image was bisected, thereby giving an estimate of the centroid or position of the image in one coordinate direction. Fortunately, an automatic measuring facility had been added to the machine as there were 60 stars in their parallax programme.

Various other measuring machines have been constructed to measure photographic plates, including the GALAXY, COSMOS and SuperCOSMOS machines at the Royal Observatory, Edinburgh. SuperCOSMOS (Hambly et al., 1998) takes

2.5 hours to digitise a Schmidt plate from the UK Schmidt Telescope and produces 2 Gbytes of data per $6^o \times 6^o$ plate, a testament to the storage capacity of the photographic emulsion.

There are two factors to consider when attempting to measure the parallax of an object. The first is to ensure that there are sufficient reference stars within the field of view of the camera with which to construct a reference frame. The bare minimum is around 4 widely separated references, with the target near the centre of the detector. The greater the number of reference stars, the more accurately a reference frame can be constructed.

The second consideration is to ensure that there is sufficient signal to noise in the images to allow accurate centroiding. For photographic plates, this requires long integrations for faint sources. The increase in sensitivity of CCD detectors over photographic plates makes parallax observations on faint sources more feasible.

In the early 1980's, large–format semi–conductor charge–coupled devices (CCD's) started to become available. One of the first parallax programmes to use these devices was on the Kitt Peak National Observatory (Monet and Dahn, 1983). This programme used a Fairchild CCD–211 with 190×244 pixels (Marcus et al., 1979).

CCD's have two distinct advantages over photography. Firstly, their detective quantum efficiency can approach 80% in some parts of the spectrum and can be

at least 40% over a large region of the spectrum (Monet, 1988). Secondly, the data from the observations are already digitised, allowing immediate observer inspection and processing of the data.

However, sampling an optical image of a star of width $\sim$1.0″ requires at least 2 pixels, fixing the pixel scale at $\sim$0.5″/pixel. This gives a field of view with early CCD's of $\sim$ 2′×2′, tiny when compared to the $\sim$ 24′×24′ field of view of the Yerkes Observatory Programme (van Altena, 1971). Initial CCD astrometry was therefore restricted to crowded fields.

Later CCD programmes have employed larger chips, *e.g.* 800×800 pixels with pixel scale of 0.206 ″/pixel giving greater sampling of images and $\sim$ 2.7′×2.7′ field of view (Monet et al., 1992) and 1024×1024 pixels with pixel scale of 0.372″/pixel giving $\sim$ 6.3′×6.3′ field of view (Tinney et al., 1995). The first programme has shown that it is possible to achieve better than 1 milli–arcsecond (*mas*) accuracy in parallax determination using optical CCD's. The second has shown that it is practical to use common user instruments on a telescope not dedicated to astrometry.

All images of the heavens are distorted by atmospheric refraction of the starlight when observing away from the zenith. If stars of different intrinsic colour are being observed, the light will be refracted different amounts, leading to a shift in the apparent relative positions of the stars — differential colour refraction (DCR). This has consequences for optical parallax work which are discussed in

more detail in section 2.7.1.

Much parallax work has focussed on dwarf red stars. To be visible from the solar system, such objects must be close and should therefore exhibit a significant parallax. A measured parallax for the object yields the distance and hence allows a good estimate of the absolute luminosity. The colour of the object coupled with the absolute luminosity, provides a strong diagnostic tool for determining the character of an object.

Much work has occurred in this area. Van Biesbroek searched for faint companions to 650 stars with known high proper motion and parallax (van Biesbroek, 1961). His search produced 12 companions. Although at that time it was not possible to measure the parallax of the faint companion directly, the matching proper motion of the faint companion allowed van Biesbroek to assign the same parallax as the bright companion. One of these objects, vB 10, is a subject of the observations described later in this chapter.

More recent work has allowed extension of the main sequence for red dwarfs right to the limit of sustainable hydrogen burning a stellar mass of 0.08 $\mathcal{M}_{\odot}$ (Henry and McCarthy, 1993; Tinney et al., 1995). Henry and McCarthey use a similar technique to Van Biesbroek. Their sample of objects are faint, red binary stars of known parallax, many taken from *The Third Catalogue of Nearby Stars* (Gliese and Jahreiss, 1991), the best map we have of our local galactic environment. Measurements of the orbits of these binaries allow an estimation to be made of

the mass, enabling a mass–luminosity relation to be established. Tinney et al. use their parallax data to construct an absolute magnitude with colour relation (M_K vs. (I–K)).

Red dwarf stars are most luminous in the infrared part of the spectrum, so the extension to the infrared of traditional tools of astronomy is a natural progression. Leggett and Hawkins have published two–colour diagrams showing a distinct main sequence in infrared colours (Leggett and Hawkins, 1988).

Brown dwarfs are extremely interesting objects expected to have mass in the range $[\sim 0.001\mathcal{M}_\odot, 0.08\mathcal{M}_\odot]$. The central temperature of these objects at formation is not high enough to sustain hydrogen burning and also leads to a low intrinsic luminosity. These objects are also most visible in the infrared. An accurate distance to such an object would yield an estimate of the intrinsic luminosity, thereby helping to determine the nature of a candidate.

Photographic techniques can explore the shortest wavelength infrared regime (*e.g.* i–band: $\lambda_{cen} \approx 7950$ Å, $\delta\lambda \approx 1250$ Å), but photographic emulsion is not sensitive at longer wavelengths ($> 1\ \mu m$). IR detectors were originally single element devices to measure the "flux in the bucket" over the collecting area. Array detectors were introduced on the United Kingdom Infrared Telescope (UKIRT) in the mid 1980's in the shape of IRCAM (McLean et al., 1986).

The array at the heart of IRCAM was 62×58 pixels, with the option of 0.6, 1.2

and 2.4″/pixel, with only the 0.6″/pixel option capable of sampling the image on what is now considered a night of poor seeing. This gave a field of view of $\sim 0.6' \times 0.6'$. The current incarnation, IRCAM3, boasts a 256×256 pixel array, with a nominal plate scale of 0.286″/pixel giving a field of view of $\sim 1.2' \times 1.2'$. This plate scale fully samples the seeing disk on UKIRT and IRCAM3 brings the usable field of view for IR astrometry closer to that available for optical CCD astrometry in the early 1980's.

The work presented here describes the first infrared observing programme aimed at testing whether infrared parallax measurements provide accurate measurements. If so, infrared observations can complement visible observations, as extremely red objects such as brown dwarfs are more easily observed in the infrared.

The observations described here were taken with IRCAM3. An significant benefit of observing in the infrared is that the effect of differential colour refraction is much reduced — section 2.7.1.

Infrared starlight is also distorted less by atmospheric turbulence allowing sharper images to be formed, *e.g.* $\lambda = 0.55\mu m, \Omega = 0.45''$, $\lambda = 2.2\mu m, \Omega = 0.34''$, where Ω is the image full width at half maximum (Chapter 3, section 3.2). The smaller field of view using IRCAM3 compared with Tinney et al. (1995) results in $\sim \frac{1}{27}$ of the area of the sky and hence number of potential reference stars being imaged. Assuming that the accuracy with which the reference frame can be constructed is described by Poisson statistics of the number of stars used to construct the

reference frame, the smaller field of view means that the reference frame will be constructed with $\sim \frac{1}{5}$ the accuracy.

The sharper images will somewhat offset the loss of accuracy from the reduced number of reference stars. Tinney et al. (1995) measured the parallax of vB 10 with an accuracy of 5.8 *mas*, so a similar observing programme in the infrared should have an accuracy of about 20 *mas* when determining the parallax of vB 10.

Soon to be installed at UKIRT is the UKIRT Fast Track Imager (UFTI) which will take advantage of the superior images available with UKIRT since the installation of the tip–tilt system. UFTI is built around a 1024×1024 infrared array with a fixed pixel scale of $0.090''$/pixel. It will be complemented by a more versatile 1024×1024 infrared array–based combined imager and spectrometer (UIST), which is currently in the design stage.

2.2 Astrometry

2.2.1 Proper Motion

The proper motion of a star is its motion with respect to the Sun projected onto the plane of the sky (Green, 1985). A star which is close to the Sun will have, on average, a high proper motion against the "fixed" background of extremely distant stars. In figure 2.1, X has equatorial coordinates (α, δ), X' has coordinates

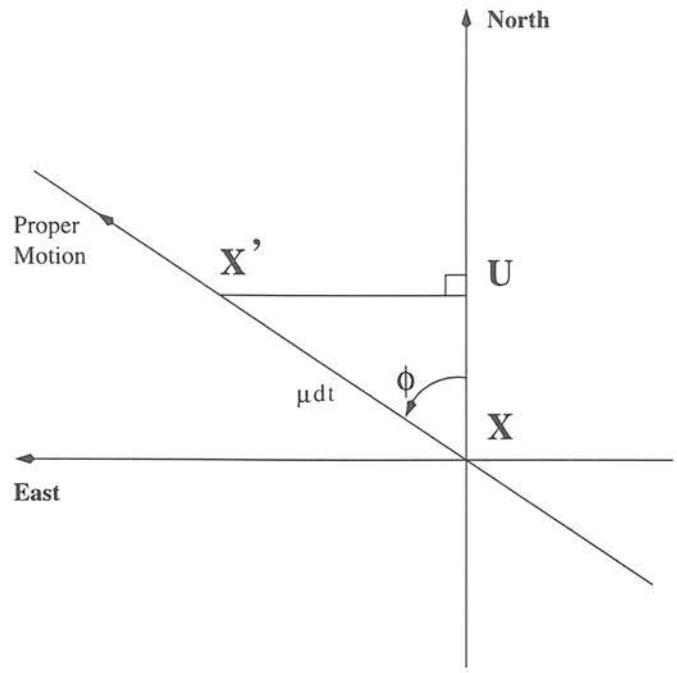


Figure 2.1: Proper motion.

$(\alpha + d\alpha, \delta + d\delta)$ so that

$$UX' = d\alpha \cos(\delta + d\delta), \tag{2.1}$$

$$UX = d\delta. \tag{2.2}$$

XX' is the distance moved by a star with proper motion μ in time dt with position angle ϕ. If the displacement is small, then UXX' can be approximated by a plane triangle with a right–angle at U, leading to the following:

$$d\alpha \cos\delta = \mu dt \sin\phi, \tag{2.3}$$

$$d\delta = \mu dt \cos\phi. \tag{2.4}$$

The components of the proper motion $\underline{\mu} = (\frac{\mathrm{d}\alpha}{\mathrm{d}t}, \frac{\mathrm{d}\delta}{\mathrm{d}t})$ are directly measurable through observations and are defined as

$$\mu_\alpha \equiv \frac{\mathrm{d}\alpha}{\mathrm{d}t} = \mu \sin\phi \sec\delta, \tag{2.5}$$

$$\mu_\delta \equiv \frac{\mathrm{d}\delta}{\mathrm{d}t} = \mu \cos\phi. \tag{2.6}$$

Proper motion is usually quoted in polar fashion (μ, ϕ) with the position angle measured from North through East, as shown in figure 2.1. The components are

$$\mu = \frac{\mu_\delta}{\cos\phi}, \tag{2.7}$$

$$\phi = \tan^{-1}\left(\frac{\mu_\alpha}{\mu_\delta \sec\delta}\right). \tag{2.8}$$

2.2.2 Trigonometric Parallax

Annual parallax is the apparent motion of a nearby star against the fixed background caused by the Earth's motion around the Sun. The geometry of the problem is shown in figure 2.2. It is obvious that

$$\underline{r} = \underline{r}' + \underline{R}, \tag{2.9}$$

where $\underline{R}$ is the position vector of the Earth in its orbit about the barycentre of the solar system. The components of $\underline{R}$ are listed in astronomical units at daily intervals in the *Astronomical Almanac*. The coordinate system is rectangular equatorial from the mean equator and equinox for the standard epoch of 2000.0.

Defining the unit vectors $\underline{s}$, $\underline{s}'$ and $\underline{s}_0$ in the directions of $\underline{r}$, $\underline{r}'$ and $\underline{R}$ respectively so that

$$\underline{r} = r\underline{s}, \quad \underline{r}' = r'\underline{s}' \quad \text{and} \quad \underline{R} = R\underline{s}_0, \tag{2.10}$$

means that equation 2.9 can be rewritten as

$$r'\underline{s}' = r\underline{s} - R\underline{s}_0. \tag{2.11}$$

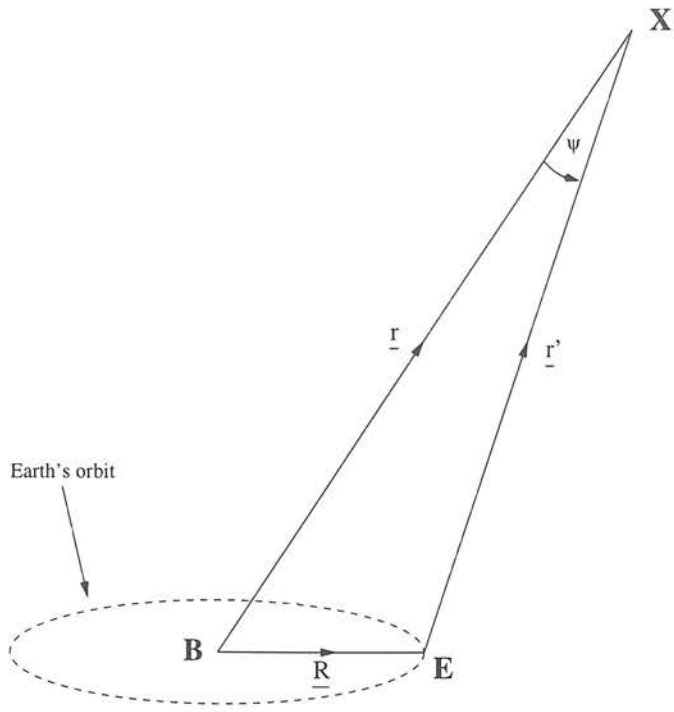


Figure 2.2: Annual stellar parallax.

Taking the vector product with $\underline{s}$ twice yields

$$r'\,[\underline{s}\times\underline{s}\times\underline{s}'\,]=\underline{s}\times\underline{s}\times[r\underline{s}-R\underline{s}_0]\,, \tag{2.12}$$

which becomes

$$\underline{s}(\underline{s}.\underline{s}')-\underline{s}'=\frac{-R\underline{s}\times(\underline{s}\times\underline{s}_0)}{r'}. \tag{2.13}$$

If $R \ll r$, then the parallactic displacement is very small giving $r' = r$ and $(\underline{s}.\underline{s}_0) = 1$, leading to the approximate equation:

$$d\underline{s}=\underline{s}'-\underline{s}=\frac{R\underline{s}\times(\underline{s}\times\underline{s}_0)}{r}. \tag{2.14}$$

Referring once again to figure 2.2, the angle between vectors $\underline{r}$ and $\underline{r}'$ is called the *annual parallax* and is denoted here as ψ. Defining $E = \mathrm{B\hat{E}X}$ and using the sine rule gives

$$\sin\psi=\frac{R}{r}\sin E. \tag{2.15}$$

The *stellar parallax* is defined as π, leading to

$$\sin \pi = \frac{1}{r}, \tag{2.16}$$

if r is expressed in astronomical units, $R = 1$ and $E = 90^o$, which corresponds to a star being on the line which passes through the barycentre, and is perpendicular to the plane of the solar system.

As the parallax of any star is small, equation 2.16 simplifies to

$$\pi = \frac{1}{r}. \tag{2.17}$$

$R = 1$ has already been adopted, so if π is expressed in radians, r is measured in astronomical units (AU). If π is quoted in arcseconds, r is expressed in *parsecs* where

$$1\ parsec\ =\ 206\ 265\ AU. \tag{2.18}$$

Substituting equation 2.17 into equation 2.14 and setting $R = 1$ gives

$$d\underline{s} = \underline{s}' - \underline{s} = \pi[\underline{s} \times (\underline{s} \times \underline{s}_0)] \tag{2.19}$$

which becomes

$$d\underline{s} = \pi[(\underline{R}.\underline{s})\underline{s} - \underline{R}]. \tag{2.20}$$

From their definitions,

$$\begin{aligned} \underline{R} &= (X, Y, Z), \text{ and} \\ \underline{s} &= (\cos\alpha\ \cos\delta,\ \sin\alpha\ \cos\delta,\ \sin\delta), \end{aligned} \tag{2.21}$$

leading to the following equations for each component of equation 2.20:

$$-\sin\alpha\,\cos\delta\,d\alpha - \cos\alpha\,\sin\delta\,d\delta = \pi[(\underline{R}.\underline{s})\,\cos\delta\,\cos\alpha - X], \quad (2.22)$$

$$\cos\alpha\,\cos\delta\,d\alpha - \sin\alpha\,\sin\delta\,d\delta = \pi[(\underline{R}.\underline{s})\,\cos\delta\,\sin\alpha - Y], \quad (2.23)$$

$$\cos\delta\,d\delta = \pi[(\underline{R}.\underline{s})\,\sin\delta - Z]. \quad (2.24)$$

Eliminating $d\delta$ with the combination [(2.22)×sin α - (2.23)×cos α] solves for $d\alpha$, whilst equation 2.24 solves for $d\delta$ directly:

$$d\alpha = \pi F_\alpha, \text{ where } F_\alpha \equiv \sec\delta\,(X\sin\alpha - Y\cos\alpha), \quad (2.25)$$

$$d\delta = \pi F_\delta, \text{ where } F_\delta \equiv (X\cos\alpha\,\sin\delta + Y\sin\alpha\,\sin\delta - Z\cos\delta). \quad (2.26)$$

2.2.3 Combined Solution

Assuming a star is close enough to exhibit both proper motion and parallax, the form of the star's apparent motion on background stars is

$$\alpha = \alpha_0 + \mu_\alpha t + \pi F_\alpha, \quad (2.27)$$

$$\delta = \delta_0 + \mu_\delta t + \pi F_\delta, \quad (2.28)$$

where t is measured in years, defined as zero when the star is at some reference position (α_0, δ_0). The units of the components of proper motion are commonly milli–arcseconds per year ($mas\ yr^{-1}$) with parallax measured in milli–arcseconds (mas). The equations have three unknowns, and therefore require at least three separate measurements for a determination of the unknowns. The parallax solution is then fitted to the points using a least–squares approach from Numerical Recipes (Press et al., 1988).

The parallax factor F_α is greater than F_δ, resulting in greater displacement in the α direction. For this reason, parallax is usually determined using only equation 2.27, although the full form of equation 2.28 is required for the proper motion fit.

Care must be taken in scheduling observations, as observations at close to zero parallax factor will not allow a determination of any parallax the object may exhibit.

2.2.4 Annual Aberration

Annual aberration is the displacement of the apparent position of a star from its true position due to the relative velocity of the Earth and the light from the object as the observation is made — see figure 2.3 which describes the classical approximation of aberration. Whilst making an observation, the velocity of the Earth ($\underline{v}$) combines with the velocity of the light ($\underline{c}$) so that the light arrives at the telescope along $\underline{v} - \underline{c}$. The Earth's velocity is approximately 30 $kms^{-1} \approx 10^{-4}c$, resulting in an angle of deviation (θ) of approximately 10^{-4} $radians$ or about $20''$.

The field of view of IRCAM3 is $\sim 75''$ allowing the approximation that the beams of light from the objects in the field are very close to parallel. It is therefore a good approximation that the displacement of the stars due to aberration is uniform across the field, *i.e.* the inter–object displacement is zero. In the data analysis stage described in section 2.6, the positions of the images of reference

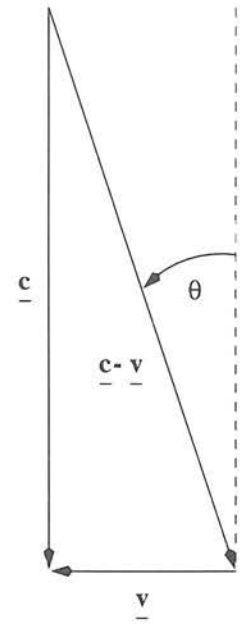


Figure 2.3: Classical approximation of annual aberration.

stars are translated in (x, y) between epochs, removing the first–order effects of aberration.

2.2.5 Correction to Absolute Parallax

The quantities measured by the combined solution in section 2.2.3 are measured relative to the reference stars. For this reason the estimate of the parallax is described as the *relative parallax*, and denoted π_{rel}. If the reference stars have any parallax of their own, the relative parallax will be corrupted, making the distance estimate inaccurate. For this reason, it is necessary to determine the average parallax of the reference stars, $\overline{\pi}_{ref}$, and correct to *absolute parallax*, π_{abs}, through

$$\pi_{abs} = \pi_{rel} + \overline{\pi}_{ref}. \tag{2.29}$$

The absolute parallax measurement can then be used to more accurately estimate the distance to an object. The following photometric technique describes a method to estimate the average parallax of the reference stars, and is taken from Tinney (Tinney, 1993).

$(R - i)$ colour is determined for each reference star, and an estimate for the absolute magnitude in the i–band, M_i, is found through the following relation (Leggett and Hawkins, 1988):

$$M_i \;=\; 4.484 \;+\; 3.116 \times (R-i) \;+\; 0.04270 \times (R-i)^2. \qquad (2.30)$$

Absolute magnitude is the brightness one would observe if the source were at 10 *parsecs*. The actual observed magnitude, m_i, can be combined with the estimated absolute magnitude to estimate the distance, and hence parallax, to each reference star through:

$$M_i \;=\; m_i \;- \log\left(\frac{d}{10pc}\right). \qquad (2.31)$$

The mean of the parallaxes of the reference stars gives the required $\overline{\pi}_{ref}$.

2.3 Observations

The observations presented in this thesis were made using the K–band (2.2 μm) filter of the IRCAM3 infrared camera on the United Kingdom Infra–Red Telescope (UKIRT) on Mauna Kea, Hawai'i. In direct imaging mode (no magnifier), IRCAM3 has a nominal pixel scale of 0.286″/pixel over a 256 × 256 array, giving a field of view of $\sim 1.2' \times 1.2'$.

Object	α B1950	δ B1950	Observations
PC 0025+0447	00 25 07.3	+04 47 08.8	8/94, 12/94, 7/95, 7/96
vB 10	19 14 32	+05 04.7	8/94, 7/95, 12/95, 7/96, 11/96, 04/97
HD 214 280	22 34 10.04	+45 47 08.6	8/94, 12/94, 12/95, 7/96

Table 2.1: Details of the observations.

Scattered across the IRCAM3 array are numerous bad pixels. To ensure that objects were not continually sitting on bad pixels, the telescope was jittered on the sky between integrations to displace the positions of objects on the array. The magnitude of this jitter ranged between 5″ and 15″. More care was taken with the observations of PC 0025+0447 as most of the potential guide stars are near the edge of the field — figure 2.5.

The jitter pattern also makes possible the use of the observations themselves in the construction of an appropriate flat field in the data reduction stage — section 2.4.

2.3.1 vB 10

As mentioned earlier, vB 10 (figure 2.4) was discovered by Van Biesbroek though a survey of 650 high proper motion stars with measured parallax (van Biesbroek, 1961). His goal was to find companion stars with faint intrinsic luminosities. Observations of the same field were performed with an interval of 5 years, and this was sufficient to detect 12 companion objects. The plates from the two

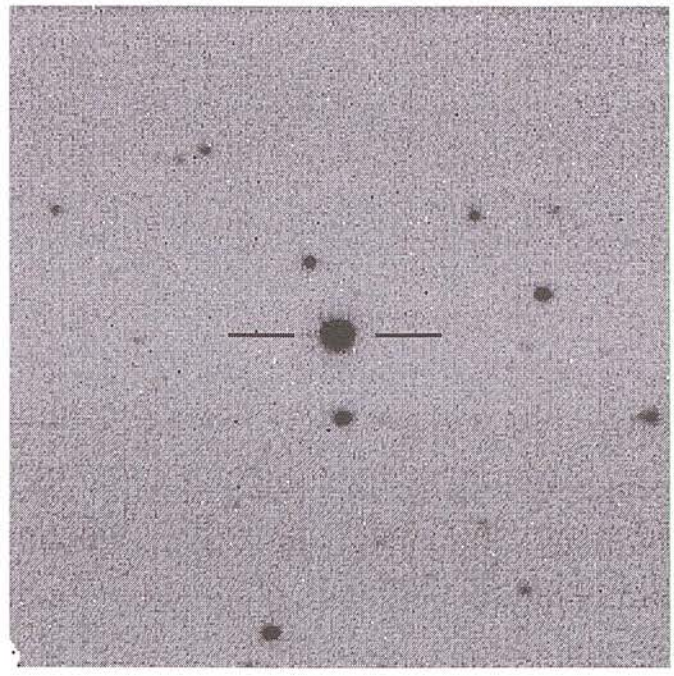


Figure 2.4: K–band finding chart for vB 10.

epochs were compared using a blink comparator.

Redder and fainter objects have since been discovered, *e.g.* PC 0025+0447 (section 2.3.2) which is the main target object in the astrometry programme. vB 10 is the archetypal late-type M dwarf and as such is very well studied photometrically and spectroscopically (Jones et al., 1994; Linsky et al., 1995) and has a well measured parallax for checking our method (Harrington et al., 1983; Monet and Dahn, 1983; Monet et al., 1992; Tinney, 1996).

The separation between vB 10 and PC 0025+0447 is about 5 hours in right ascension, meaning that when PC 0025+0447 is near maximum parallax factor, vB 10 is near minimum. As our observations were optimally scheduled to observe PC 0025+0447, service observations of vB 10 were therefore performed at

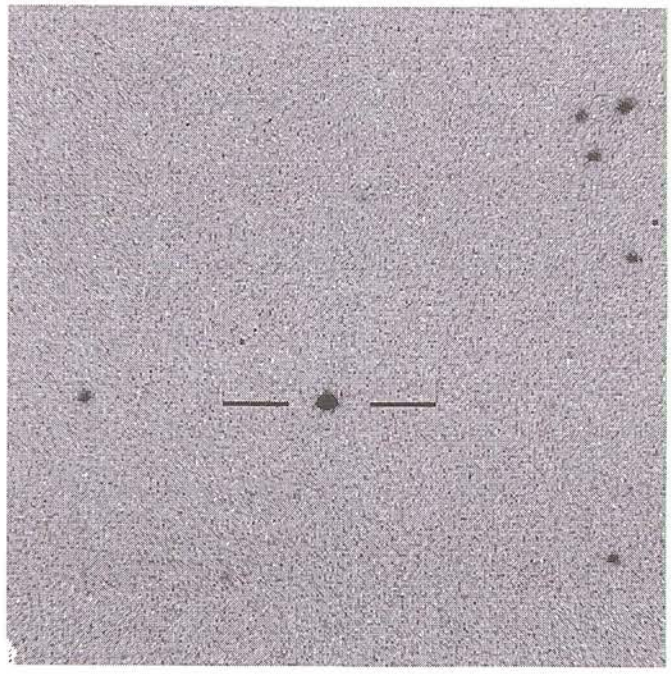


Figure 2.5: K–band finding chart for PC 0025+0447.

2 further epochs to observe vB 10 at near maximum parallax factor.

2.3.2 PC 0025+0447

PC 0025+0447 (figure 2.5) is an extremely red object discovered as part of a grism survey for high redshift quasars (Schneider et al., 1991). It was detected by an automated search algorithm due to its very strong (>250 Å equivalent width) H alpha emission line. Subsequent spectroscopic and photometric follow-ups have classified it as type dM9.5e, and it appears to be a cooler, fainter analogue of LHS 2924 — a field brown dwarf candidate (Graham et al., 1992; Kirkpatrick et al., 1993; Schneider et al., 1993; Mould et al., 1994).

A good estimate of the distance to PC 0025+0447 would be of great interest.

Finding charts suggested that there would be enough objects within the IRCAM3 field of view to act as suitable references for an astrometric investigation of PC 0025+0447. To this end, PC 0025+0447 was observed at 4 epochs as detailed in table 2.1. The observations were scheduled to observe PC 0025+0447 at near maximum parallax factor. At each epoch, measurements were made on two consecutive half–nights. On each half–night, data for PC 0025+0447 was taken for at least 2 hours.

2.3.3 HD 214 280

Additional observations were made of the triple star system HD 214 280. This system was selected from the HIPPARCOS input catalogue where it has numbers 111582, 111584 and 111580 for components A, B and C respectively (ESA, 1997). This system was chosen so that the behaviour of the focal plane of UKIRT could be investigated. The extremely accurate positional measurements expected from the HIPPARCOS astrometry mission would also allow an independent estimation of the plate scale of IRCAM3.

Figure 2.6 shows the field. The output catalogue of HIPPARCOS reports that all three components share the same parallax and proper motion, therefore their relative positions are unchanged through time. Of the three components, only two, A and C, are visible in the K–band. The arrow on figure 2.6 shows the approximate location of component B. The nature of object X is unknown, but

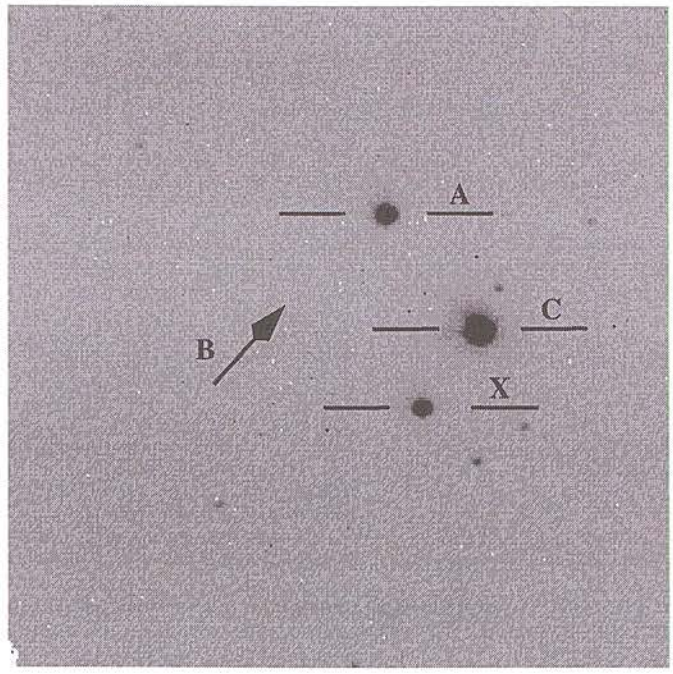


Figure 2.6: K–band finding chart for HD 214 280.

its brightness at K and proximity to the two visible components of HD 214 280 make it useful.

The observing method for HD 214 280 was slightly unusual. To investigate whether there are any significant plate scale variations across the IRCAM3 array, the system was imaged at 25 locations in the array using a 5×5 mosaic with a separation between locations of $15''$ in both α and δ. The aim of these observations was to provide enough resolution to measure any gross plate–scale variations.

2.4 Data Reduction

Instrumental and thermal effects mean that the images obtained with IRCAM3 are not immediately suitable for analysis. The raw images obtained from IRCAM3, $R(\underline{r})$, are a combination of contributions from various sources described by the following equation:

$$R(\underline{r}) = g(\underline{r}) \times [I(\underline{r}) + s] + D(\underline{r}), \tag{2.32}$$

where $I(\underline{r})$ is the image of the sky we wish to recover, s is the *sky background* light level which should be a scalar over such a small area of sky, $g(\underline{r})$ is the gain of each pixel of the detector — commonly called the *flat field*, and $D(\underline{r})$ is the *dark current* — the signal obtained in the absence of light for the same integration time.

Rearranging equation 2.32 gives:

$$I(\underline{r}) = \left[\frac{DS(\underline{r})}{g(\underline{r})}\right] - s, \tag{2.33}$$

where

$$DS(\underline{r}) = R(\underline{r}) - D(\underline{r}) \tag{2.34}$$

are the dark subtracted frames. Equation 2.33 specifies the data reduction scheme, which is implemented with some additions in standard packages such as the routine `stred` in the standard IRCAM reduction package, IRCAMDR. Here is a general outline:

1. All raw images are scaled to a common integration time.

2. The dark current, $D(\underline{r})$, is subtracted.

3. The detector response is removed by dividing by the flat field, $g(\underline{r})$.

4. The images should be scaled to a common airmass as the atmosphere has an empirically measured extinction.

5. The sky background, s is subtracted leaving a mean background level of zero signal.

It is common to amalgamate the final step into the image analysis stage of the process. To obtain a well–reduced image, the dark current and the flat field must be obtained in addition to the integration time and extinction which are known.

The data reduction described below was implemented in the IDL programming environment (RSI, 1994), making use of the ASTROLIB public domain library routines. The method closely follows the standard method used in the routine `stred` in IRCAMDR.

2.4.1 Dark Current

The dark current can readily be measured during a series of observations by performing an integration with the shutter of the camera closed. It is common practise to include a dark integration as part of a set of observations on a particular object and this was done throughout our observations. The dark frame

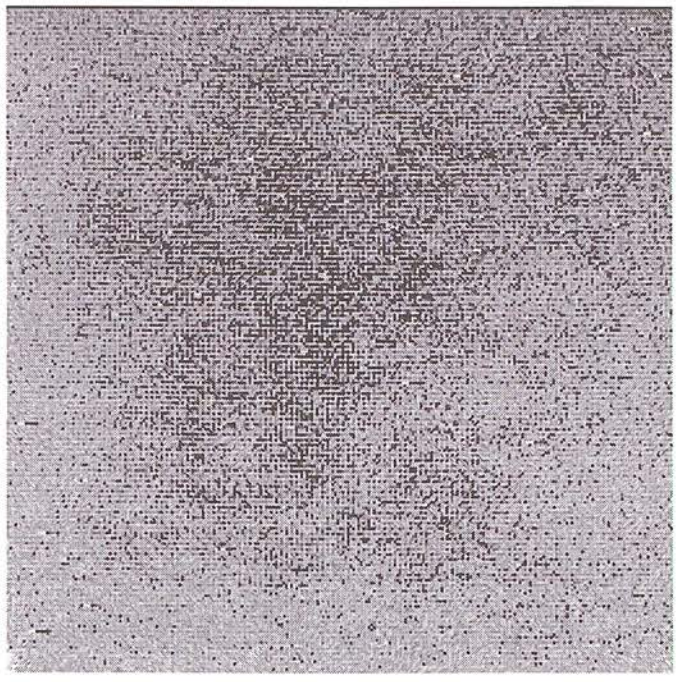


Figure 2.7: A typical dark frame from IRCAM3 — scaling is linear in the range $[\mu - 3\sigma$ (black), $\mu + 3\sigma$ (white)], where $\mu = 15.25$ and $\sigma = 0.38\mu$.

taken at the beginning of each set of observations is used to correct the following integrations of the sky.

The dark is independent of the filter being used for the observations as it is a measure of the charge which builds up through thermal and other effects in addition to charge that collects due to incident photons.

2.4.2 Flat Field

The flat field, $g(\underline{r})$, is the spatial response of the detector to uniform illumination. Figure 2.8 shows a typical flat for IRCAM3. There are characteristic stripes across the whole array which have, possibly apocryphally, been attributed to the wipe

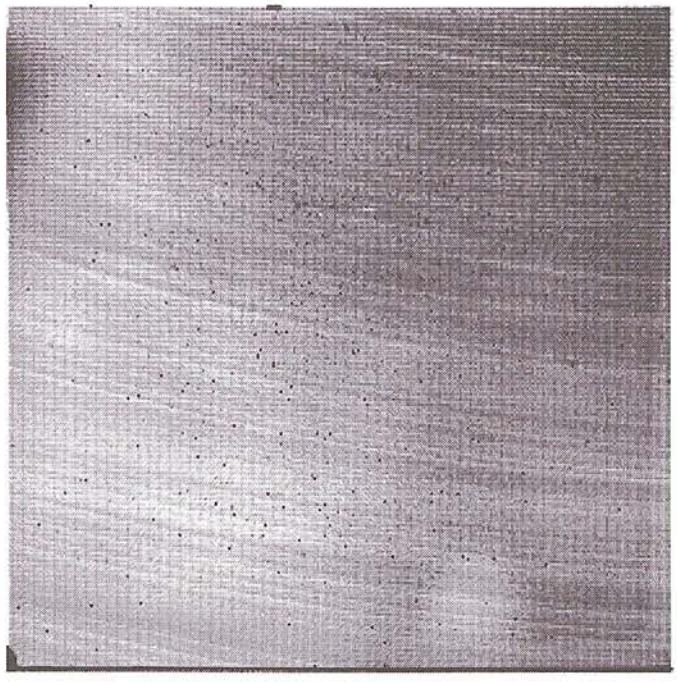


Figure 2.8: A typical flat field for IRCAM3 — scaling is linear in the range $[\mu - 3\sigma$ (black), $\mu + 3\sigma$ (white)], where $\mu = 0.9981$ and $\sigma = 0.0498\mu$.

of a cloth across the surface of the chip. There is also a "splodge", bottom middle–right, which rumour suggests is a fingerprint.

Construction of a flat field requires a series of observations of a fairly empty piece of sky. The pointing of the telescope is altered between observations in a raster pattern to ensure that the array views different areas of the sky. This jittering is essential for rejection of sources which could contaminate the flat field. A dark–subtracted dataset is built up of the form:

$$\text{Sky raster} = DS(\underline{r}, k), \tag{2.35}$$

where k indexes the raster pattern of the observations. The flat field is obtained from the sky raster by taking each position in $\underline{r}$ and determining the median value

from the k samples at that position. In mathematical notation:

$$m(\underline{r}) = \mathrm{med}_k[DS(\underline{r}, k)]. \tag{2.36}$$

$g(\underline{r})$ is obtained by scaling m($\underline{r}$) to a median of 1.0 to preserve the correct intensity scale in equation 2.33:

$$g(\underline{r}) = \frac{m(\underline{r})}{\mathrm{med}_{\underline{r}}[m(\underline{r})]} \ . \tag{2.37}$$

There are several caveats which must be applied to equation 2.36:

1. This has already been mentioned: the field used to construct the sky raster must be of a reasonably empty piece of sky. This is to ensure that most of the data points used to construct the median are measurements of the sky and not a source which would contaminate the flat field.

2. The raster pattern must have a large enough throw so that any objects (stars) in the field do not occupy the same position on the array too often. Again, this is to get as many sky measurements per median determination as possible.

3. A large number of frames to construct the raster is preferable, again to increase the number of measurements of the sky in the median determination.

4. Equation 2.36 assumes that the sky brightness is uniform and constant over the sky raster and also that the response of the detector is constant.

Experience has shown that while items 1 to 3 are easy to satisfy, item 4 is not.

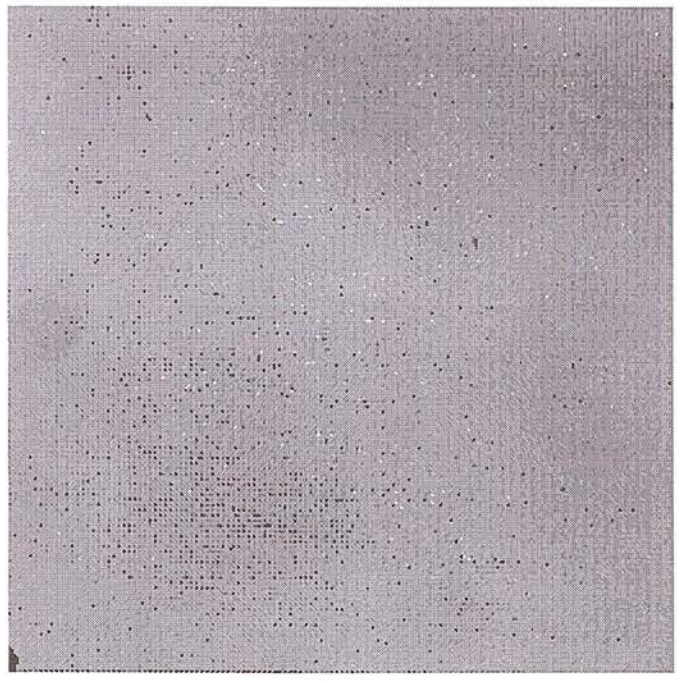


Figure 2.9: A comparison of two flat fields constructed from data taken on the same night — scaling is linear in the range $[\mu - 3\sigma$ (black), $\mu + 3\sigma$ (white)], where $\mu = 1.0003$ and $\sigma = 0.0017\mu$.

Figure 2.9 shows the ratio of two flat fields from the same night, constructed from consecutive sets of sky rasters. Whilst the gross stripes and splodges have been correctly identified in both cases and are missing from the ratio, it is clear that there is variation in the flat field through time.

There are two quite clear trends in figure 2.9. Firstly, there are a significant number of individual pixels whose response has varied significantly — the isolated white/black pixels. Secondly, there are broad variations in response of order 100 pixels or so.

It is quite clear that the individual pixel variation is intrinsic to the detector, but not so clear from one comparison whether the broad variation is caused by sky

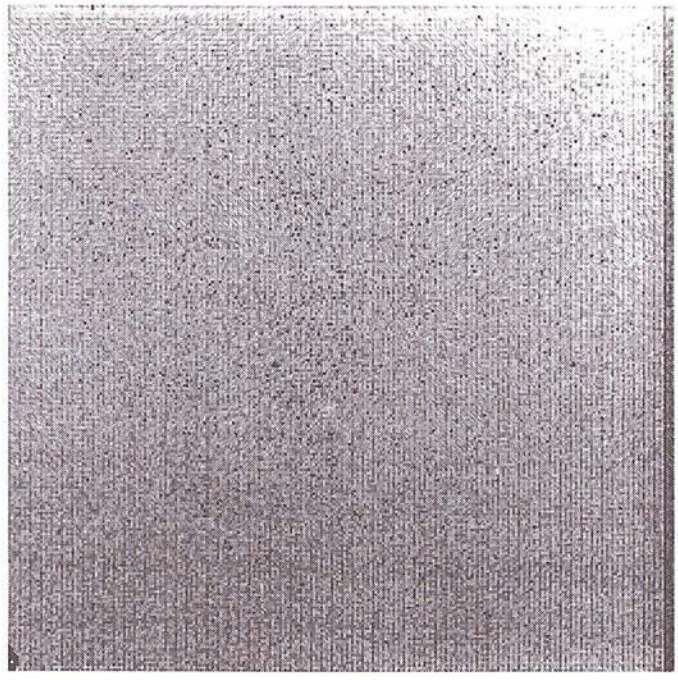


Figure 2.10: A reduced frame of PC 0025+0447 using a sky–produced flat field. The observations used to construct the flat field were made about 30 minutes before this integration on the object — scaling is linear in the range [$\mu - 3\sigma$ (black), $\mu + 3\sigma$ (white)], where $\mu = 645$ and $\sigma = 0.06\mu$.

background or detector variations. Comparing many different flat fields yields similarly shaped broad distortions regardless of epoch of observation, so it is safe to assume that the broad features are also instrumental effects.

The variation of the response of the instrument with time places quite strict constraints on how the flat field is created, especially in when reducing the PC 0025+0447 observations. The reduction process for these observations will now be discussed in detail.

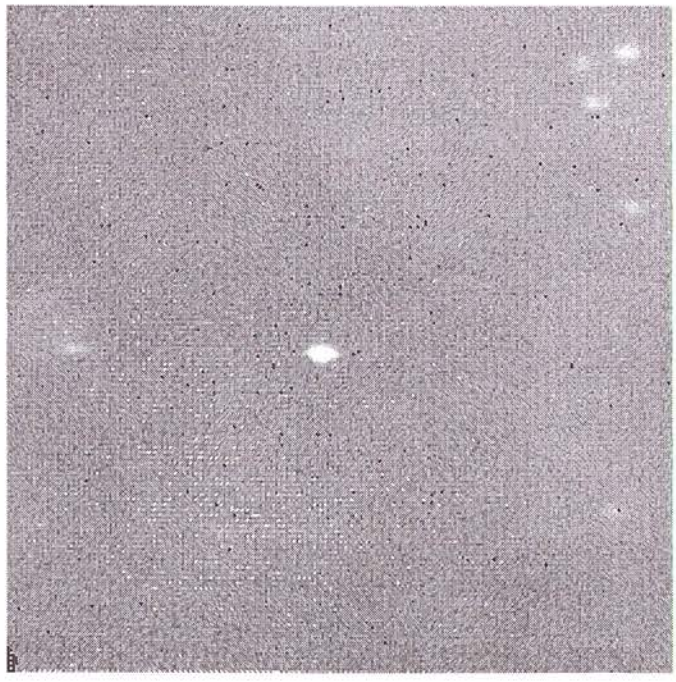


Figure 2.11: A reduced frame of PC 0025+0447 using a median–filtered flat field constructed from too many ($\sim$ 30) object frames — the distortions apparent are similar to those in figure 2.9 — scaling is linear in the range [$\mu - 3\sigma$ (black), $\mu + 3\sigma$ (white)].

2.4.3 Reducing the PC 0025+0447 observations

Sky Frames

The most straightforward method of constructing a flat field is to use a set of sky raster frames of a random patch of sky. This was done to create figure 2.10 where the raster was observed about 30 minutes before the observation of the object was taken. It is quite clear that the flat field is hopelessly wrong — there is very little useful information to be obtained from such an image.

Many Object Frames

Using the observations of the object is a perfectly valid method of creating a flat field, provided the caveats outlined in section 2.4.2 are followed. Figure 2.11 shows an image of PC 0025+0447, this time using a flat constructed from a large number of frames of PC 0025+0447. The same irregularities as were found in the comparison between flat fields (figure 2.9) can again be seen in the background of this image, so it is clear that the flat field used in this instance is incorrect. The variations imply that the frames used in the construction of this flat field span too long in time, as the response of the detector has varied whilst the observations were being made.

Few Object Frames

Using a smaller number of frames to construct the flat field does not give the instrument long enough to vary its response significantly. Figure 2.12 shows the same frame as in figure 2.10, but this time reduced using a flat field constructed from a group of 8 frames which includes the displayed observation.

In general, the background is flat which shows that the flat field is accurate. However, the relatively small number of frames used to construct the flat field has led to breaking caveats 1 to 3 in section 2.4.2. This has led to artefacts of the stellar images being present in the flat field.

Figure 2.12: A reduced frame of PC 0025+0447 using a median–filtered flat field constructed from an appropriate number of frames to suppress the flat field variations — scaling is linear in the range [$\mu - 3\sigma$ (black), $\mu + 3\sigma$ (white)], where $\mu = 647$ and $\sigma = 0.0018\mu$.

The presence of stellar images in the sky stack which is median filtered displaces the medians upwards for the regions of the array which are contaminated. As the reduced image is obtained by division by the flat field, the slightly higher stellar artefacts in the flat field cause "holes" to appear in the reduced images. These holes betray the raster pattern of the telescope when making the observations.

The artefacts are most easily seen around PC 0025+0447 as it is the brightest object in the field. This image of PC 0025+0447 is in the centre of the raster, and careful study of the image in figure 2.12 shows 4 dark artefacts surrounding the bright image, forming the five spots on a die.

The image of PC 0025+0447 in this frame will also be sitting on top of an arte-

fact in the flat field, as this image was one of those used to construct the flat. Additionally, the three closely grouped stars in the top–right of the frame will be affected, not only by artefacts of themselves, but also by artefacts of the other stars as the separations of the stars are similar to the throw used by the telescope to create the raster pattern.

Modified Few Object Frames

There is something of a problem here in that a good estimate for the gross flat field requires only a small number of frames in the sky raster, while removal of stellar artefacts requires a large number. Any artefacts in the flat field will distort the stellar images, an undesirable effect regardless of the subsequent image processing performed. In the image analysis described later, the major effect of the artefacts is to increase the scatter on the estimate of the centroids of stellar images.

A novel iterative process was developed in the course of this work which aims to remove the stellar artefacts from the few object frame flat fields. The improved method uses the DAOPHOT application to detect and subtract stars in reduced images. DAOPHOT is also used in the image analysis stage to detect and estimate positions (centroids) of stars in the reduced frames and is described in more detail in section 2.5.

The method constructs the initial flat field in the same way as that in figure 2.12

and divides the dark subtracted frames, $DS(\underline{r})$, by this flat in the normal way to get reduced frames in which the artefacts are apparent. This gives a set of frames, $I_i(\underline{r})$, to which we apply DAOPHOT to remove the stellar images, $S_i(\underline{r})$. If all stellar images are removed, this results in an artefact corrupted image of the sky only:

$$I'_i(\underline{r}) = I_i(\underline{r}) - S_i(\underline{r}), \tag{2.38}$$

where the prime ($'$) denotes removal of stellar images.

The corrupted sky images are then multiplied by the existing flat which removes the artefacts in the flat fielding as these were caused in the division by $g(\underline{r})$ to construct the images $I_i(\underline{r})$ in the first place. Star subtracted, dark subtracted frames are obtained:

$$DS'_i(\underline{r}) = g(\underline{r}) \times I'_i(\underline{r}). \tag{2.39}$$

There may be errors in the subtraction of the stars in the DAOPHOT process, but these will be small compared with the addition to the sky brightness level from having a star in the image, so the aim of the process of removing the cause of the corruption of the median away from the true value will be largely successful.

It is possible that through stellar images sitting on bad pixels and other effects, not all stars in a frame are detected and subtracted. This reduces the efficacy of the technique, though it is still worthwhile as the scatter of centroid estimates is reduced.

The star subtracted, dark subtracted frames are used to construct an artefact–

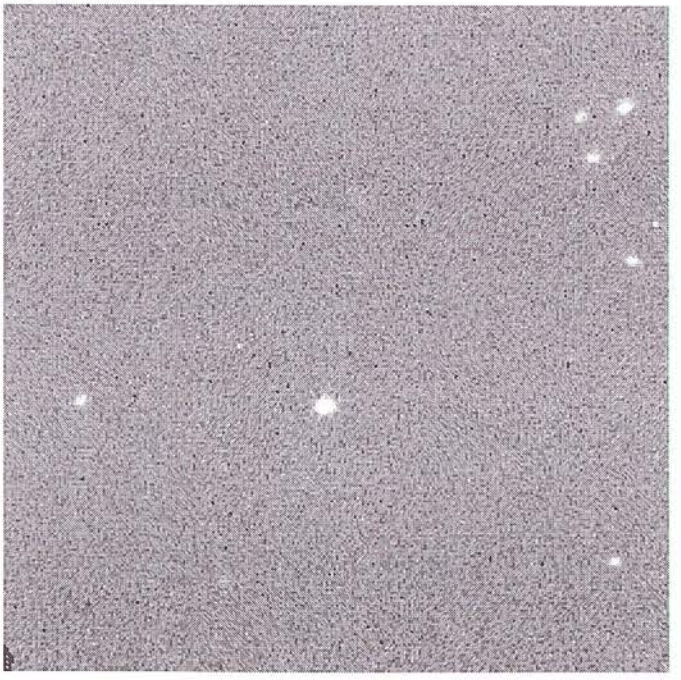


Figure 2.13: A reduced frame of PC 0025+0447 using the artefact–reduced flat field — scaling is linear in the range $[\mu - 3\sigma$ (black), $\mu + 3\sigma$ (white)], where $\mu = 647$ and $\sigma = 0.0017\mu$.

reduced flat field:

$$g'(\underline{r}) = \frac{\mathrm{med}_i[DS'_i(\underline{r})]}{\mathrm{med}_{\underline{r}}[\mathrm{med}_i[DS'_i(\underline{r})]]} \ , \tag{2.40}$$

which is then used to reduce the original dark subtracted frames, $DS(\underline{r})$, to produce artefact reduced final images.

It is possible that the reduction of the artefacts enables a star previously undetected by DAOPHOT to become detectable. The above process is implemented iteratively, until the number of stars detected falls below one per frame. This gives the best star subtracted, dark subtracted images from which to construct the final flat field.

Figure 2.13 shows the same frame as in figure 2.12, but reduced using the artefact–

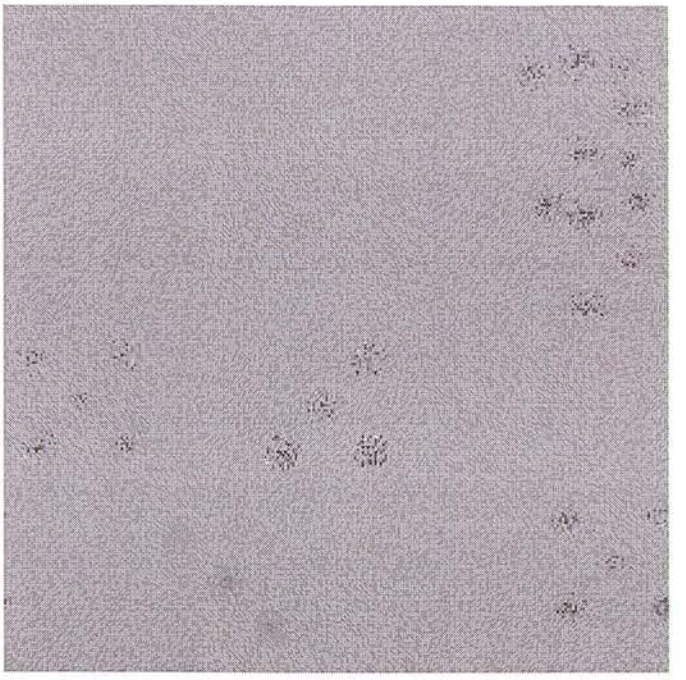


Figure 2.14: Difference between figures 2.12 and 2.13 — scaling is linear in the range [$\mu - 10\sigma$ (black), $\mu + 10\sigma$ (white)], where $\mu = -0.0110$ and $\sigma = 0.0785$.

reduced flat field method. Most of the artefacts around PC 0025+0447 have been suppressed. Perhaps more enlightening is figure 2.14 which is the difference between figures 2.12 and 2.13. This shows the holes which have been removed by the improved reduction technique.

Removal of the holes using this technique results in an increase in positional accuracy. The standard error on the position for PC 0025+0447 within a night reduces to about $\frac{2}{3}$ of its value without the iterative technique.

2.4.4 Reducing the vB 10 observations

The reduction of the frames of vB 10 was performed using separate sky frames to create the flat field. This was done for two reasons:

1. The number of stars in the field is larger than the PC 0025+0447 field — some of the observations go quite deep with > 50 objects in the field.

2. The objects in the field are significantly brighter and therefore cover a larger number of pixels than those in the PC 0025+0447 field.

These effects conspire to make median filtering of the object images impossible as there are not enough pixels in which just the sky is exposed. This forces the use of separate sky frames to construct the flat field. There will be errors in the flat fielding which will affect the centroiding of the stars, which will be offset by an increase in accuracy through centroiding on the brighter objects in the vB 10 field.

2.4.5 Reducing the HD 214 280 observations

The HD 214 280 observations were reduced using the "many object frames" method. The short integration time employed per frame coupled with the large number of frames making up a set of observations allows this method to be employed successfully.

2.4.6 Airmass Correction

Correction of the images to common airmass allows for easier cross identification of the stars in the data analysis stage — section 2.6. The IRCAMDR package lists the extinction due to the atmosphere for the K–band filter used in IRCAM3 at UKIRT as 0.080 magnitudes per airmass (Leggett, 1996). To scale to a common airmass, images are modified in the following manner:

$$I_{amcor}(\underline{r}) = I(\underline{r}) \times 10^{\left(\frac{0.080 \times airmass}{2.5}\right)}. \tag{2.41}$$

This is applied to the flat fielded images before they go on to the image processing stage.

2.5 Image Processing: DAOPHOT

DAOPHOT is an astronomical application for obtaining precise photometry and astrometry for stars in digital images and was developed to be able to perform crowded field photometry where blending of stellar images is a significant problem (Stetson, 1987). To perform this function, DAOPHOT is able to empirically determine the shape of the point spread function (PSF) from objects within a field. This PSF is then fitted to objects within the field, yielding accurate positions and magnitudes for all the detected objects. In the fitting process, DAOPHOT groups stars which are likely to be blended and performs simultaneous fits for all stars in the group.

DAOPHOT is available as part of the ASTROLIB libraries for IDL, and these routines were used as part of an IDL application to perform the image processing. A typical run of DAOPHOT is performed by using the following routines (DAOPHOT routines are in **bold**):

1. **SKY**

   Sky finds the mean and 1-σ width of the sky brightness histogram for about 4000 pixels throughout the frame.

2. **FIND**

   This routine searches for objects in the frame which might be stars. It uses a threshold method based on the values obtained by SKY to flag candidates. The routine requires an estimate of the full width at half maximum of an unresolved stellar image to help select only stars. It then performs two tests on these features before an object is confirmed as a star.

   The SHARP and ROUND tests are designed to reject cosmic rays and other spurious candidates from the list of potential stars. The default acceptance regions for the tests have proved very reliable, with any significant deviation from the defaults resulting in numerous artefacts being selected as stars. Approximate positions and magnitudes are returned for each star.

3. SORT

   The candidate stars from FIND are ordered in increasing magnitude so that the brightest is first.

4. **APER**

Aperture photometry is performed on the candidate stars for a user specified range of aperture radii which are centred on the position of each candidate star. An annulus of pixels is also used within which a determination is made for the local sky brightness enabling a more accurate estimation of the magnitude of the star to be estimated.

APER detects and rejects any candidate stellar images which contain bad pixels, fall off the edge of the frame or are holes caused by errors in the flat field.

5. **GETPSF**

GETPSF finds the empirical PSF for the field from the list of candidate stars. The initial PSF is constructed from the first in the list, with contributions from the other objects added in with their relative contributions weighted with reference to the magnitudes found by APER.

The user specifies an approximate FWHM and size of the array into which the PSF will be recorded. This is then decomposed into an analytic Gaussian profile which is a fit to the core of the PSF and a table of residuals to the Gaussian fit to correct the analytic PSF to the empirically obtained PSF. These two sets of information are written directly to file by GETPSF so it is not forgotten.

6. **RDPSF**

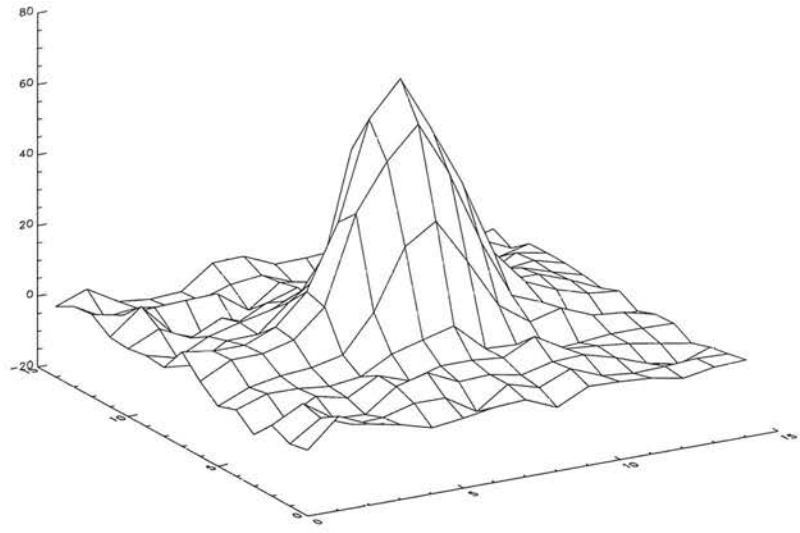


Figure 2.15: A PSF obtained by DAOPHOT.

This routine reads the appropriate PSF datafiles and reconstructs the PSF in an appropriately sized array.

7. SHOWPSF

The PSF is displayed as a perspective surface plot for visual inspection as shown in figure 2.15. This is the PSF found for the objects in figure 2.16 — it is reassuring that the shape of the PSF correctly reflects the elongation in the right ascension direction due to the known RA wobble on UKIRT.

8. **GROUP**

For crowded field photometry, DAOPHOT needs to know which stars are likely to be blended. This is determined at this stage by comparing the separation of the stars found with FIND and the size of the array upon which the PSF is determined. If, whilst fitting to one star, another will

influence the fit, then both stars are placed into the same group so that their fit can be performed together.

9. **NSTAR**

   The routine which performs the final fit. If more than one star is in a group, then all stars for the group are fitted simultaneously. The previously estimated magnitudes and centroids are used as initial estimates for the iterative routine.

   The output of this routine is written to file and lists the magnitude and centroid of the stellar image, along with the SHARP and ROUND coefficients obtained in the fit. These data files, one per observation, are used as the input for the astrometry software which is described in section 2.6.

10. READSTARS

    This routine reads the data file written by NSTAR.

11. **SUBSTAR**

    The stars fitted by NSTAR are removed from the original image creating a star–subtracted frame.

12. DISPLAY

    Plots original (figure 2.16) and star–subtracted (figure 2.17) images for comparison.

At the end of the image processing stage, there is a DAOPHOT data file per

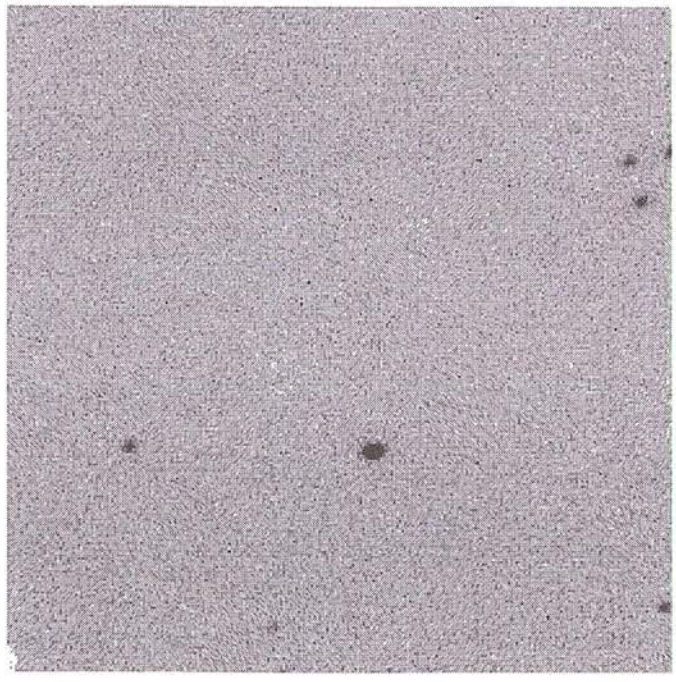


Figure 2.16: A reduced image before DAOPHOT processing — scaling is linear in the range [$\mu - 3\sigma$ (black), $\mu + 3\sigma$ (white)].

Figure 2.17: A reduced image after DAOPHOT processing where the detected stars have been subtracted — scaling is the same as in figure 2.16.

reduced image for use in the astrometric analysis.

2.6 Astrometry Software

The astrometry software has been written to be as general as possible and has two parts. Part one takes the many individual DAOPHOT output files of an object of interest for one night of data and constructs a local astrometric solution, the *night master*. Part two aligns all the night masters for that object to a global reference frame, detects stars which exhibit any motion and performs a parallax and proper motion fit for these stars. The programmes will now be described in more detail.

2.6.1 Nightly Astrometric Solutions

The entries in the DAOPHOT datafile which are of interest here are the coordinates and magnitudes of the stars which have been detected in each IRCAM3 image. The initial origin of the coordinate system for each file is the bottom left corner of the IRCAM3 array. The job of this part of the astrometry software is to define and align all the files to a common coordinate system. The general outline of the process is sketched in figure 2.18.

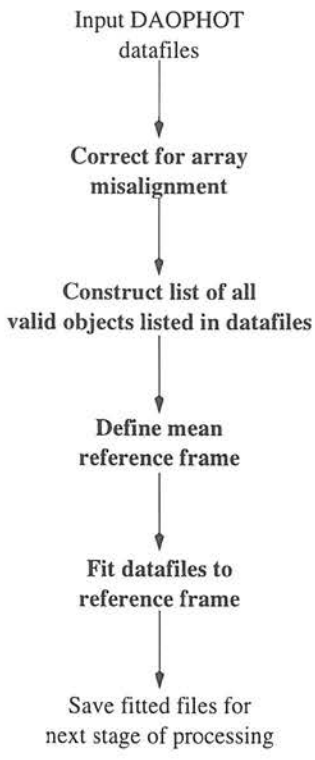


Figure 2.18: The process by which the nightly reference frame is defined.

Preprocessing

As described in section 2.7.3, North of the IRCAM3 array is misaligned 1.8° Eastwards with respect to North ($\delta = +90°$). The first operation of the programme is to remove this rotation. There is then an option to compensate for any spatial magnification effects which might need to be removed. As described later in section 2.7.2, removing any spatial magnification effects derived from the HD 214 280 observations did not improve the subsequent fit, so this option is disabled in normal operation.

Preliminary Master List

To be able to define a reference frame from all the objects found by DAOPHOT, it is necessary to construct an exhaustive list of all distinct objects listed in the datafiles. Due to bad pixels and the jitter pattern of the observations, all stars will not be listed in all datafiles. The datafile with the greatest number of stars is selected as a preliminary master list against which the other files are compared.

Exhaustive Master List: Coordinate Matching

Each datafile is compared in turn to the master list. Before comparison can take place, the two lists must share similar coordinate origins. This matching of origins is performed in the following manner. The datafile origin is transformed to its brightest star, and then the origin of the master list is transformed to each star in the master list in turn. For each master file origin, the stars whose positions match within a tolerance are totalled. The origin in the master list with the most matches will correspond to the same origin in the datafile, thereby matching the coordinate systems.

Exhaustive Master List: Object Comparison

Comparison of objects in the datafile and master list is by magnitude in the first instance. If there is dubiety within a user defined tolerance, then the positions of the candidates in the matched coordinate systems are compared. If comparison of

the positions does not resolve the dubiety, then the dubious stars in the datafile are ignored. Due to the jittering of the telescope, there will be stars found in datafiles which are not in the original master list. If this is the case, then they are added to the master list. Thus, the master list will grow with time as each datafile is compared.

Once all the files have been compared, the master list will contain details of all objects which have been found by DAOPHOT in the image processing for the night's observations. It is possible that there are artefacts in this list which are only found in one file. The number of datafiles in which an object in the master list is found is determined and those objects with only 1 occurrence are deleted from the master list. The master list now contains details of all valid objects found by DAOPHOT in field of interest for the night in question.

Comparing Datafiles With Master List

Copies of the master list are made, one for each datafile. The datafiles are compared to the master list, and the corresponding entries in the copy of the master list for which there are measurements in the datafile are replaced with the values listed in the datafile. The other entries in the copy of the master list, for which there is no corresponding data in the datafile, are replaced with a universal BAD DATA flag.

At the end of this comparison, there is a modified copy of the master list for each

datafile, with the only real entries in a copy being the data that was determined by DAOPHOT in the corresponding datafile — the positions listed are in the local DAOPHOT datafile coordinate system. All other entries in the copies have the BAD DATA flag. Selecting an entry in the master list, one can then inspect each modified copy and determine in which datafiles that object was detected.

Choice of Local Coordinate System

The modified copies must now be expressed in a common coordinate system. The brightest object in the master list is chosen as the origin of the coordinate system and the coordinates of the modified copies are transformed appropriately. The mean and its standard error are determined for each star by averaging over the set of modified copies, discounting BAD DATA entries. The mean of the brightest star has been explicitly defined as zero, with an undefined standard error by its selection as the origin of the coordinate systems.

An iterative least–squares fit is then performed aligning each modified copy in both translation, (x, y), and magnification and rotation (m, θ) onto the frame defined by the mean positions of all the stars. The fitting iterates until the summed standard error stops decreasing.

When the fitting has converged, the modified copies have each been transformed to the night master reference frame. The mean position of the brightest star will no longer be exactly zero, and there will be a defined standard error on this mean, a useful check that the fitting routine has performed correctly. The fitted modified copies are then saved to file, completing this stage of the data processing.

2.6.2 Global Astrometric Solutions

The scheme employed to determine a global astrometric solution for any stars which exhibit motion is shown in figure 2.19. The datafiles saved by each night master stage are loaded and the mean positions and standard errors on the mean determined, defining each separate night master frame. These frames must be expressed in a common coordinate system, and a similar process to that described above is performed to cross match stars and choose a preliminary origin for the coordinate system. A crucial difference is that only those stars which appear in all night master frames are selected for use in determining the reference frame.

Once the selection of stars and provisional choice of origin has been made, the processing continues in iterative fashion. A weighted least–squares fit defines a provisional global reference frame. At this stage, the reference frame could be corrupted by the presence of stars with high parallax and/or proper motion. Can-

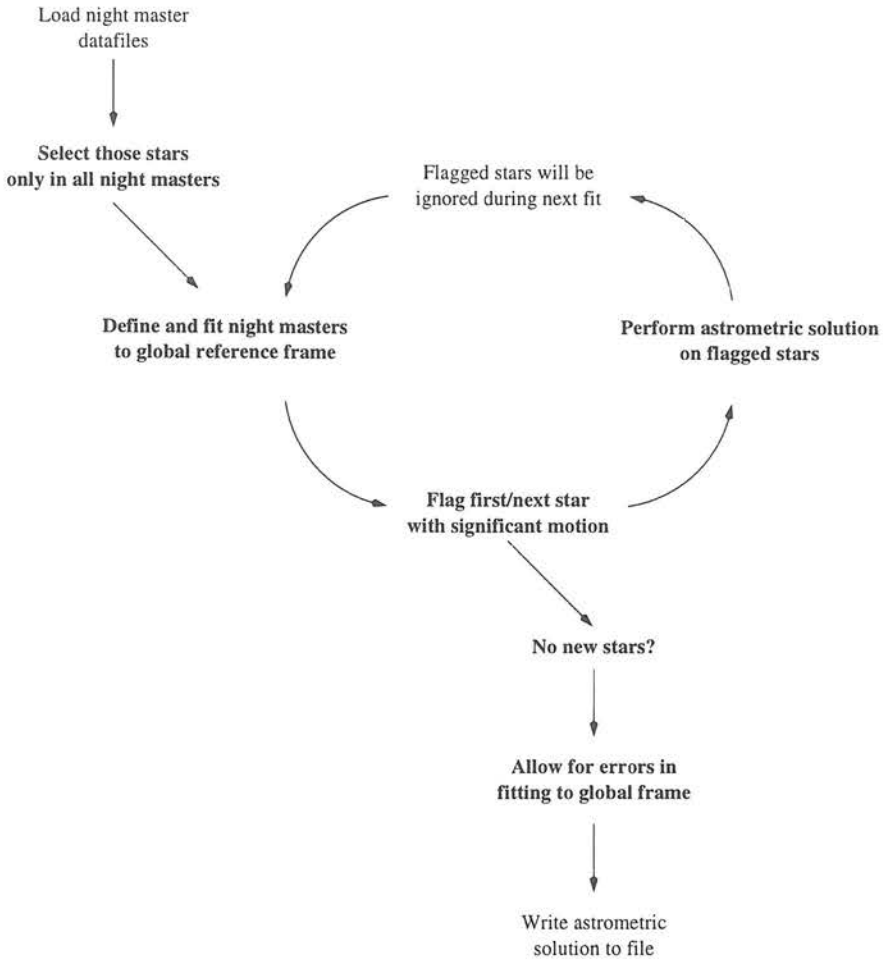


Figure 2.19: The process by which the global reference frame is defined and the astrometric solution performed.

didate parallax and proper motion stars are found by performing a preliminary astrometric solution against the provisional reference frame.

The star exhibiting the largest parallax is flagged and the night masters are refitted to the global reference frame ignoring the parallax star, thereby allowing the reference frame to be refined by removing the influence of the parallax star. A refined astrometric solution is then performed for the candidate and reference stars.

If one of the reference stars exhibits significant parallax, then it too has corrupted the fitting and the flagging and refitting is reperformed for both this star and the first. This process is repeated for all stars which exhibit significant motion.

Once the list of candidate stars has been exhausted, the remaining stars define the global reference frame and the candidate stars have their solution performed against this reference.

The fitting of the night master frames to the global reference frame will almost certainly contain some error. This must be accounted for if the error in the fitting is not to corrupt the final parallax solutions. To quantify the error in fitting, an astrometric solution is performed on each of the reference stars and an average parallax and proper motion determined for the reference stars. This quantifies how well the night masters have been fitted to the global reference frame. Ideally, this mean solution should be zero, with very small errors.

The fitted night master frames are corrected by the appropriate amounts to correct for the errors in fitting to the global reference frame. The final parallax solutions are performed for the candidate stars with the results written to file.

Correction to π_{abs}

As discussed in section 2.2.5, the parallax determined above is measured relative to the background stars, π_{rel}. Appropriate colours are obtained for each refer-

ence star, enabling an average photometric parallax for the reference frame to be estimated and applied as a correction to the measured parallax to obtain an estimate for the absolute parallax, π_{abs}.

2.7 Limitations in Astrometric Precision

2.7.1 Differential Colour Refraction

Differential Colour Refraction (DCR) in astronomical imaging arises when light is non–normally incident on the atmosphere. Blue light is refracted through a greater angle than red light, causing a shift in the relative positions of stellar images. The precise magnitude of this effect depends on the spectra of stars which are imaged and the transmission characteristics of the filter.

The combination of the shapes of the stellar spectra and filter transmission result in a different effective wavelength for each star in the field. At zenith, starlight is normally incident on the atmosphere, so it is a good approximation that over the field of view of IRCAM3, there is no DCR, regardless of the wavelength of observation. However, as the field of interest moves from the zenith, the refraction of the atmosphere will affect the starlight. An object with a blue effective wavelength will be displaced towards the zenith when compared with an object with red effective wavelength. The further the field moves from the zenith, the greater the DCR effect.

DCR is accepted as the fundamental limitation of ground based optical astrometry (Monet et al., 1992; Tinney, 1996). These publications describe an empirical method for quantifying and correcting for this phenomenon based on the colour of the stars in each field. The method requires that a rich field of stars with known colours is observed at at least twice in one night. The observations should take place as far apart as possible, and on opposing sides of the meridian so that any relative motion can be attributed to DCR effects alone.

Accurate observations allow an empirical determination of DCR coefficient with object colour for a specific filter. The relation can then be used to correct for DCR effects by modifying the positions of each star in the DAOPHOT datafile by an appropriate amount determined from the airmass at which the measurement was made. This would be in the first stage of the astrometric software immediately after correcting for the misalignment of the array.

In 1996, Tinney compared the DCR coefficient for observations of vB 10 in broad (1250 Å) and narrow (90 Å) i band filters. Figure 2.20 shows the range of the DCR coefficient as listed by Allen across the i band. The range of the coefficient is reduced by a factor of ~ 13 due to the narrowing of the passband alone.

Tinney measured the DCR for the two ESO filters and determined that the effect of DCR is reduced from 11.5 to 0.1 (mas/tan Z).(r-i)$^{-1}$. Tinney concludes that for astrometry at the several milli–arcsecond level, DCR effects when using the narrow i band filter are essentially negligible. Comparison of the range of DCR

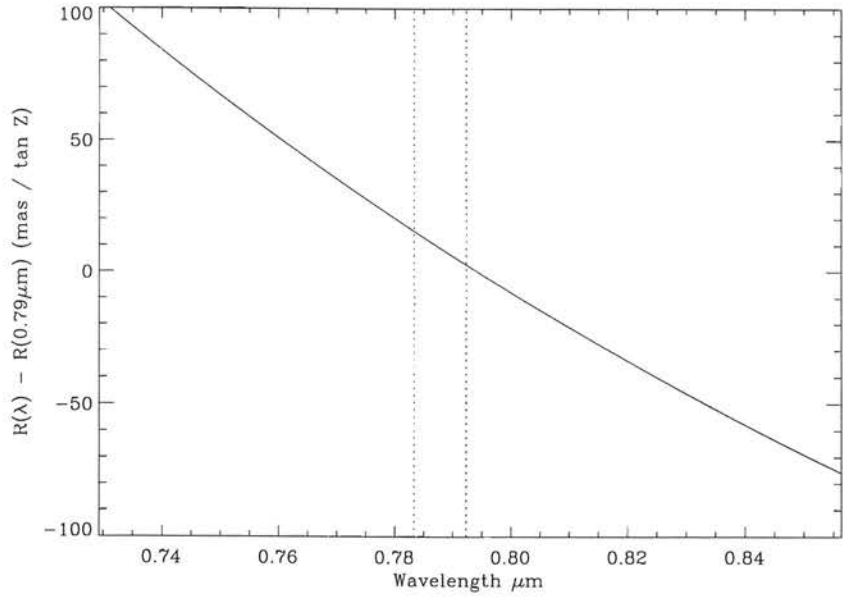


Figure 2.20: DCR coefficient across i band filter: the vertical lines show the narrow filter used by Tinney 1996 to reduce the DCR effect — data from Allen 1973.

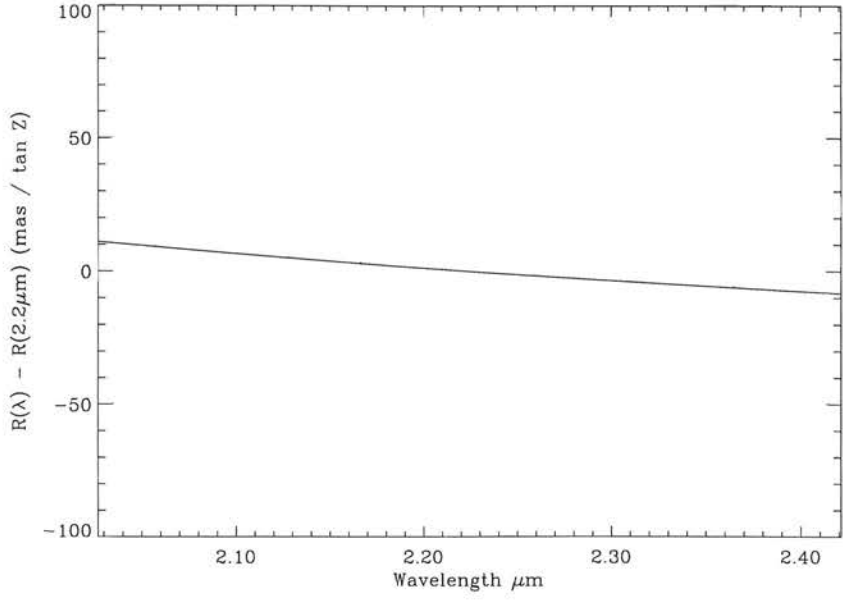


Figure 2.21: DCR coefficient across K band filter — data from Allen 1973.

coefficients in the K (figure 2.21) and narrow i bands shows a similar reduction in the DCR effect. DCR effects can therefore be ignored unless formal uncertainties in the astrometric solutions approach the milli–arcsecond level.

Attempts were made to measure the DCR coefficient using data taken in the programme. The field of vB 10 would provide an ideal rich frame with which to make the measurements, but the scheduling of the observations required that vB 10 be observed either shortly after rising or shortly before setting. PC 0025+0447 was observed over a wide range of hour angles, but the faintness of the objects in the field did not allow a successful determination to be made.

2.7.2 Focal Plane Stability

A crucial requirement for accurate astrometry is long–term stability of the focal plane arrangement. Variations of the gross plate scale between observations are compensated in the fitting of the night master frames to the global reference frame. Of interest in this section is any spatial variation of the plate scale across the IRCAM3 array.

To investigate this, observations were made of the triple star system HD 214 280. As discussed earlier, only two of the group (A & C) are actually visible in K, but a near line of sight coincident object (X) replaces the absent component in the analysis which follows. Observations were made at most epochs, placing the system at 25 places in the focal plane. Each focal plane investigation images the

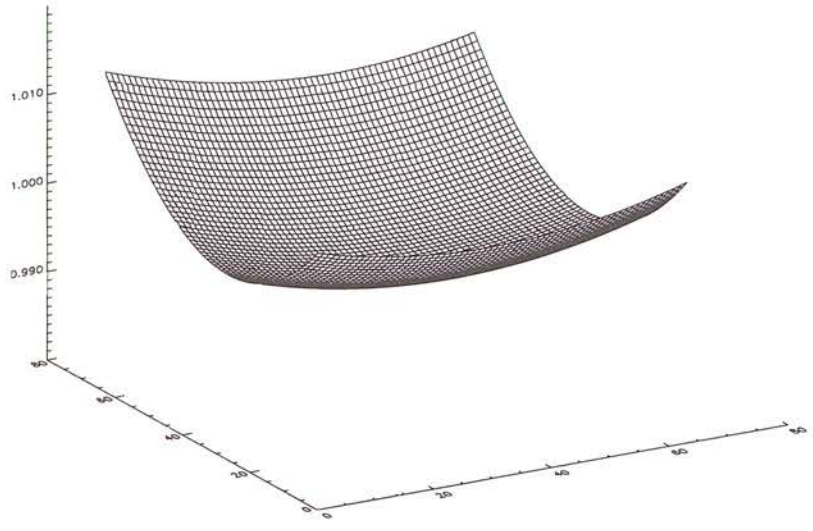


Figure 2.22: The magnification surface for August 1994. The surface is a low–order polynomial surface fit to the data–points. The 256×256 array has been resampled to 64×64 to produce a clearer surface.

three vectors between the stars at 25 different positions on the array.

For each investigation, the mean lengths of the three vectors are determined over the 25 locations on the array. The relative magnification at various locations across the array is found by dividing the length of each image of a vector by the relevant mean, assigning the location of the measurement of magnification as the mid–point of the vector giving a set of (x, y, mag) datapoints.

A 2^{nd}–order polynomial surface is then fitted to the datapoints and the function defined on a 256^2 grid corresponding to the IRCAM3 array — figure 2.22. A magnification surface is produced for each epoch with the typical standard deviation of the magnification across the empirical surface of around 0.5%.

To test whether the measured surfaces were accurate, an experiment was performed on the vB 10 data for those nights where a magnification surface was measured. The uncertainty in the position of vB 10 in each night master frame was compared to the normal reduction in two cases. Firstly, the measured focal plane variations were removed from the centroid data before being processed by the reduction scheme, and secondly, an additional effect from the measured focal plane variations was added before the reduction scheme was employed. In both cases the uncertainty in the position of vB 10 increased, showing that temporal plate scale variation could not be detected. Varying the degree of the polynomial fitted to the (x, y, mag) datapoints or substituting Zernike polynomials as the function to be fitted could not improve the accuracy of the night master positional estimates for vB 10.

A second method to allow for variable spatial plate scale was to modify the magnification part of the (m, θ) part of the least–squares fitting routine. Instead of a global magnification correction measured by the lengths of vectors between stars, the fitting was allowed to perform independent x and y magnification corrections from the corresponding components of the vectors. This modification also made matters worse when fitting vB 10 frames to a night master.

The degradation of the local astrometric solution when attempting to compensate for spatial plate scale variations leads to the conclusion that the empirical magnification surfaces are noisy and any plate scale variations are of order $< 0.5\%$.

When determining the global astrometric solutions for our objects, no correction for spatial magnification effects was made.

2.7.3 Plate Scale of UKIRT

As well as being used to attempt to measure spatial plate scale variations, the observations of HD 214 280 have allowed an independent measurement of the plate scale and misalignment of the IRCAM3 array with North. When the global reference frame is created, the night masters are fitted to it with the mean magnification and rotation over the period spanned by the observations.

The plate scale is applied when the astrometric fit is interpreted, whereas the correction for mis–alignment of the array is the very first operation in construction of the night masters. These values were originally taken from the IRCAM3 webpage, but the HD 214 280 observations allow an independent determination.

Night masters were constructed for each set of 25 observations of HD 214 280 with the correction of rotation set to zero. These contain two of the three stars in the HD 214 280 system, and comparison of the measured vector between these stars with that obtainable from the HIPPARCOS positions allows the plate scale and misalignment to be calculated. Our measurements sparsely sample time from July '94 to July '96, and are listed in table 2.2 along with other measurements made by support scientists at UKIRT.

Date (Reference)	Plate Scale ("/pixel)	Rotation ($^\circ, N \rightarrow E$)
1996	0.284	1.5
Colin Aspin	–	–
1997	0.281	1.8
Sandy Leggett	$\pm$ 0.002	$\pm$ 0.4
1994–96	0.2840	1.8
This work	$\pm$ 0.0013	$\pm$ 0.1

Table 2.2: Measurements of UKIRT plate scale and array misalignment.

The method used by UKIRT support astronomers is as follows. Two consecutive measurements are made for one rich field but with a fairly large offset of the telescope between the observations. There is some, but not complete, overlap between the stars in the two images. If there is a discrepancy between array North and true North, then there is a rotation required to fully align the objects in the two measurements. The plate scale can also be found by comparing the telescope throw with the offset observed on the chip.

Note: In the headers of the UKIRT datafiles, the recorded offsets are from the encoders and are not very accurate. If UKIRT was auto-guiding, then the pointing is to 0.1 arcsec, and the offsets (specified by the observer in an EXEC file) are accurate with the errors on the positions of the ends of the throw being the pointing tolerance. It is just the process of recording the information which is noisy (Wright, 1998).

2.8 Parallax Solutions

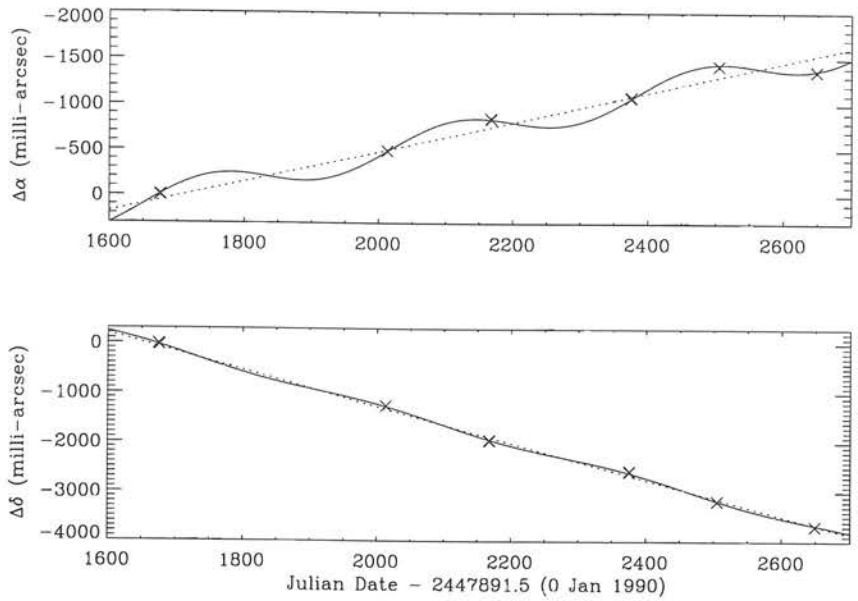
Our field of view ($\sim 75''$) is small compared with the 6.35' used by Tinney *et al.* (1995) in their recent CCD parallax programme. This presents a problem in that the number of reference stars with which one can obtain the astrometric solution is small. The vB 10 field has many field stars, whereas the PC 0025+0447 field has only 7 stars that can be used for the solution.

2.8.1 vB 10

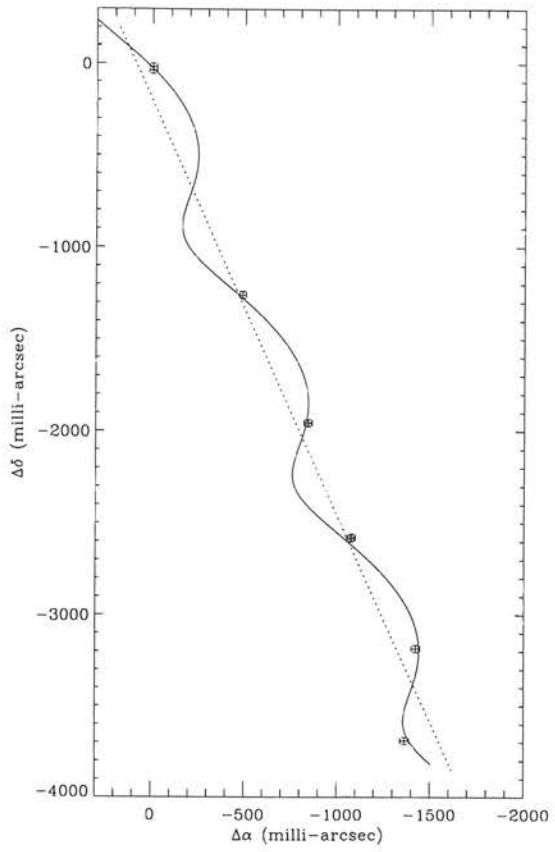
Our observations were optimised to observe PC 0025+0447 at maximum parallax factor. This meant that, except for our final two measurements, we observed vB 10 at close to zero parallax factor — Figure 2.23.

Although the number of stars in the field is large, the number of reference stars with which to find the solution is limited to the number found on the night of poorest signal–to–noise ratio. The observations in December 95, taken at air–mass 1.8, limit the number to 10. The final parallax solution for vB 10 is listed in table 2.3 along with other recent measurements of its parallax. The formal error for this solution is 16.8 *mas*, which is better than that expected given the argument in section 2.1.

The solution obtained for vB 10 is in good agreement with previous results proving the reliability of the method described earlier. The HIPPARCOS measurement



(a) The proper motion and parallax solution for vB 10.



(b) Apparent motion of vB 10 on the sky.

Figure 2.23: The proper motion and parallax solution for vB 10.

vB 10

Telescope (Reference)	D (m)	π_{rel} (mas)	μ_{rel} ($mas\ yr^{-1}$)	θ_{rel} (^{o})
USNO	1.55	173.6	1488.4	203.7
(Harrington et al., 1983)		± 4.9	± 1.1	± 0.0
Kitt Peak*	4	178.9	1483.6	204.7
(Monet and Dahn, 1983)		± 1.8	± 2.2	± 0.2
USNO	1.55	169.7	1490.3	203.27
(Monet et al., 1992)		± 0.8	± 0.8	± 0.03
Mt. Palomar	1.5	166.8	1512.9	203.3
(Tinney et al., 1995)		± 5.8	± 35	± 0.6
ESO	2.2	160.5	1479.4	202.9
(Tinney, 1996)		± 3.5	± 15.3	± 0.5
HIPPARCOS†		170.26	1452.1	203.5
(ESA, 1997)		± 1.37	± 1.0	± 0.1
UKIRT‡	3.8	162.2	1474.1	203.8
Figure 2.23		± 16.8	± 16.0	± 0.5

PC 0025+0447

Telescope (Reference)	D (m)	π_{rel} (mas)	μ_{rel} ($mas\ yr^{-1}$)	θ_{rel} (^{o})
Mt. Palomar	1.5	—	17.9	160.8
(Tinney et al., 1995)			± 7.1	± 13.1
UKIRT‡	3.8	12	11	60
Figure 2.24		± 8	± 53	± 130
UKIRT*	3.8	19	21	50
Figure 2.25		± 9	± 26	± 40

Notes:

* *Listed as a preliminary solution.*

† HIPPARCOS *measurement is for GL 752A, the companion to vB 10 (GL752B) — listed parallax is* π_{abs}.

‡ *This work.*

* *This work, with longer baseline using an additional frame supplied by Schneider.*

Table 2.3: Parallax measurements.

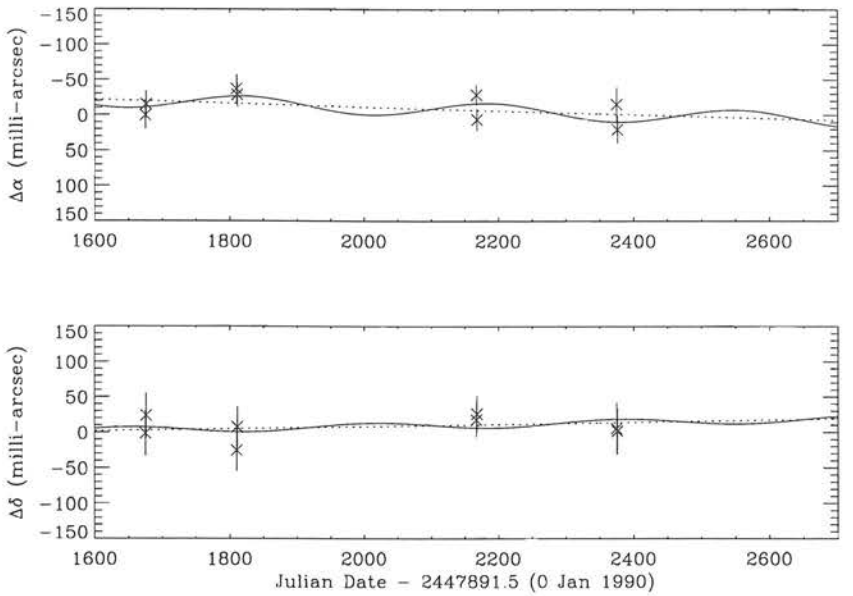
is for the companion to vB 10, GL 752A, as vB 10 is not in the HIPPARCOS catalogue. The formal errors on the astrometry solution are of order 10 milli-arcseconds, so the result is not sufficiently accurate to be significantly corrupted by DCR effects as discussed in section 2.7.1. If DCR effects were to be apparent, they would have most effect in the observations taken in December 95 due to their extremely large airmass, and the points for these two nights should be discrepant. Inspection of the fit shows that this is not the case, further evidence that DCR effects are minimal for K band observations.

Using frames kindly supplied by Tinney (private communication), $\overline{\pi}_{ref}$ was determined to be 1.7 ± 0.6 mas by the method outlined in section 2.2.5. The results in the absolute parallax of vB 10 being estimated as $\pi_{abs} = 163.9 \pm 16.8$ mas.

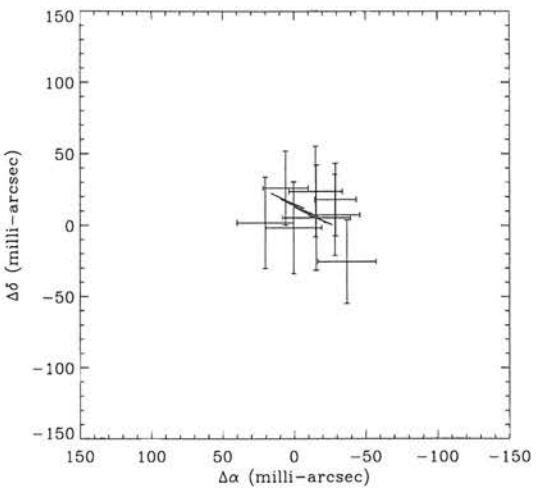
2.8.2 PC 0025+0447

The field of PC 0025+0447 is sparse, but long integrations at each epoch enable 7 other stars within the field to be used as references when determining the astrometric fit. The formal solution for PC 0025+0447 using our measurements is shown in figure 2.24 and tabulated in table 2.3. This solution estimates the parallax as $\pi_{rel} = 12 \pm 8$ mas, and places a 1–σ limit of on the proper motion of 52 $mas\ yr^{-1}$.

The data presented in Figure 2.24 are consistent with there being no measured parallax or proper motion. A solution for proper motion alone yields a marginal



(a) The proper motion and parallax solution for PC 0025+0447.



(b) Apparent motion of PC 0025+0447 on the sky.

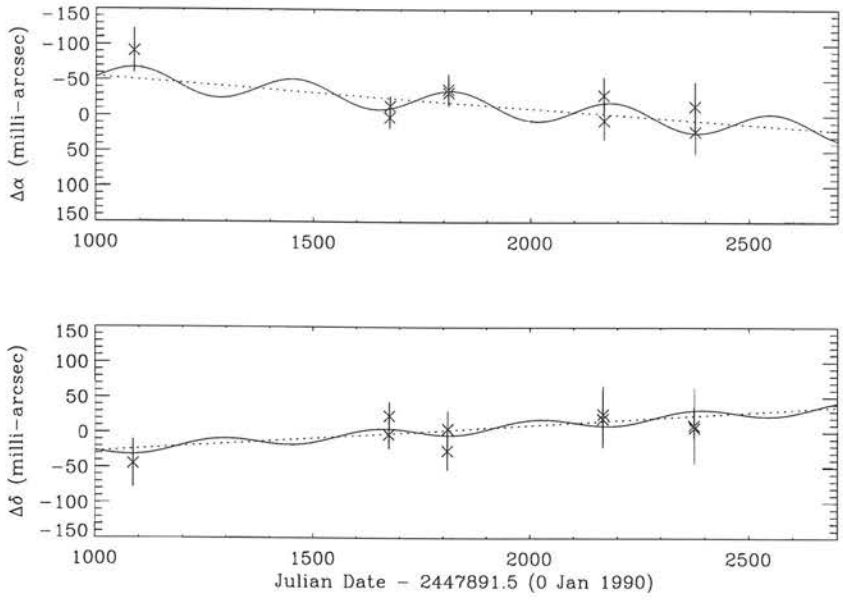
Figure 2.24: The proper motion and parallax solution for PC 0025+0447.

increase in the rms deviation of the datapoints from the fit.

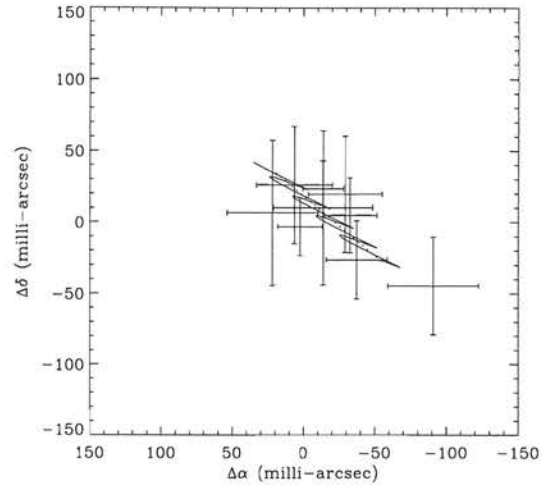
A longer baseline for the solution has been obtained by using a frame of the field kindly supplied by Donald Schneider (private communication) — figure 2.25 and table 2.3. The estimate for the parallax from this solution, $\pi_{rel} = 19 \pm 9\ mas$, is only indicative as the effects of DCR on the i–band frame have not been removed. Depending on the filter characteristics, DCR effects could move the apparent position of PC 0025+0447 with respect to the reference stars by $\sim 30\ mas$ in the RA coordinate. The long baseline proper motion solution is perhaps of more significance. It suggests a small proper motion to the North–East of $\sim 20\ mas\ yr^{-1}$. This disagrees with Tinney (Tinney et al., 1995), but agrees with Schneider (Schneider et al., 1993).

If the formal solution for the parallax from the UKIRT data is accepted, some speculation as to the nature of PC 0025+0447 is possible. Correcting to absolute parallax has little effect as $\overline{\pi}_{ref} = 0.3 \pm 0.2\ mas$. The frames used to make this estimate were kindly supplied by Schneider. The absolute parallax is found to also be $\pi_{abs} = 12 \pm 8\ mas$. This places PC 0025+0447 at 83 parsecs which, combined with $m_K = 14.95$ from Graham (Graham et al., 1992), gives $M_K = 10.4$.

This is just beyond the end of the calibration curve of absolute magnitude M_K against mass in solar units $\mathcal{M}/\mathcal{M}_\odot$ found for late-type M dwarfs by Henry and McCarthy (Henry and McCarthy, 1993). The sequence ends at $M_K \sim 9.8$ and $\mathcal{M}/\mathcal{M}_\odot \sim 0.08$, the mass limit for hydrogen burning. The absolute magnitude



(a) The proper motion and parallax solution for PC 0025+0447.



(b) Apparent motion of PC 0025+0447 on the sky.

Figure 2.25: A proper motion and parallax solution for PC 0025+0447with a longer baseline using an i–band frame supplied by Schneider.

for PC 0025+0447, $M_K = 10.4$ thus places it beyond the main sequence and into the main sequence/brown dwarf transition with an estimated mass of 0.065 $\mathcal{M}_\odot$.

2.9 Summary

In this chapter, an infrared observing programme aimed at measuring the first infrared trigonometric parallax has been described. Difficulties particular to IRCAM3, a common user imager at UKIRT were highlighted in the data reduction stage and a work–round proposed. Measurements have been performed for two objects. The measured parallax for vB 10 is in good agreement with previous work and proves the techniques which were employed. In particular, no allowance was made for differential colour refraction, as the effect in the K–band is minimal. The measured parallax for PC 0025+0447 is inconclusive, though speculation based upon the inconclusive measurements suggests that PC 0025+0447 may indeed be a brown dwarf.

Chapter 3

Adaptive Optics

3.1 Atmospheric Turbulence

The fact that the atmosphere limits the resolution of astronomical imaging has been known for a long time. In the 18^{th} century, Newton commented that,

> *"the Air through which we look upon the Stars is in perpetual tremor,"*

describing what is now known as *seeing*. He went on to suggest,

> *"The only Remedy is a most serene and quiet Air, such as may perhaps be found on the tops of the highest Mountains above the grosser Clouds"* (Newton, 1730).

Modern astronomers have followed this advice, and most large telescopes (those with a primary mirror of diameter ~4 metres or greater) can be found on top of high mountains with stable climates enabling year-round observing.

Unfortunately, this is only a partial solution. By using high mountains, astronomers can get their telescopes above most of the lower atmosphere turbulence and weather, but even at 14,000 feet above sea–level, *e.g.* on Mauna Kea, Hawaii, there is still a great deal of air above the Observatory.

Seeing is caused by movement of the atmosphere. The diurnal cycle results in differential heating of the Earth's surface which creates convection cells. These cells have many scales, including that required to produce a pretty cumulus cloud on a summer day and those responsible for the large–scale circulation of the atmosphere (Thom, 1987).

These naturally occurring motions in the atmosphere are due to changes in temperature which alter the density of pockets of air. Hotter, less dense air is more buoyant, so it rises causing motion. When air moves, such as in convection cells or in a high–altitude jetstream, there is a shear at the boundary with air moving at a different velocity. Eddy currents are generated at these shears and transfer of heat occurs, again altering the local density of the air. As the refractive index of a gas is density dependent, this also varies.

The optical path s, encountered by a ray of starlight as it passes through the atmosphere to a telescope observing at the zenith may be approximated as

$$s(\underline{r}) = \int_{h_{tel}}^{h_{atm}} n(\underline{r}, h) c \; dh, \tag{3.1}$$

where c is the speed of light, h is height above sea–level, $n(h)$ is the variation of refractive index with height and $\underline{r}$ is the position of the ray in the telescope pupil. Clearly, if there is no variation of n with $\underline{r}$, *i.e.* $n(\underline{r}, h) \rightarrow n(h)$, then the optical path encountered by all rays entering the telescope is constant and the atmosphere has had no effect other than to slow the light's progress a little. An image of a star viewed through such an atmosphere will be diffraction limited.

On the other hand, if path length variations are introduced across the pupil of the telescope then the image will be distorted. Viewed for a few seconds, images can be seen to *twinkle* – vary randomly in intensity, *quiver* – move in the focal plane of the telescope and *spread* – broaden from a diffraction limited image.

Perhaps the best way to observe this is to project an image of the sun onto a screen. Close inspection of the edge of the bright disc shows that the image is "boiling" in small, fairly well contained cells. It is possible to determine the dominant wind direction by observing where these cells move tangentially to the bright disc. Also of note is that these distortions fluctuate rapidly. The atmosphere is constantly moving and evolving.

3.2 Characterising Seeing Effects

The Kolmogorov theory of turbulence (Kolmogorov, 1941) has been shown to be a good representation of the characteristics of the atmosphere (Dainty and

Scaddan, 1975; Beckenridge, 1976; Boyd, 1978). The general premise is that eddies are set up at the largest scale, L_0, which decay into eddies of decreasing scale until frictional effects dissipate the energy at the smallest scale, l_0 (Beckers, 1993). There are however, occasions when non–Kolmogorov behaviour has been found (Nicholls et al., 1995).

A prediction of the theory (Tatarski, 1967) is that the *phase structure function*, which describes the spatial ($\underline{r}$) variation of the phase in the wavefront after transmission through the atmosphere, has the form:

$$D_\phi(r) = \langle |\phi(\rho + r) - \phi(\rho)|^2 \rangle = 6.88 \left(\frac{r}{r_0}\right)^{\frac{5}{3}} \text{ rad}^2. \tag{3.2}$$

The $r^{\frac{5}{3}}$ power law has been observed by Dainty and Scaddan (Dainty and Scaddan, 1975) and Beckenridge (Beckenridge, 1976). r_0, introduced by Fried (Fried, 1965), is the *atmospheric coherence length*, an estimation of the mean size of a turbulent cell in the atmosphere along the line of sight as described by:

$$r_0 = 0.185 \lambda^{\frac{6}{5}} \cos^{\frac{3}{5}} \zeta \left(\int_{h_{tel}}^{h_{atm}} C_N^2(h) dh \right)^{-\frac{3}{5}}. \tag{3.3}$$

$C_N^2(h)$ is the *refractive index structure constant*, which quantifies the strength of the turbulence with height. Campaign measurements of $C_N^2(h)$ characterise the properties of a location at an astronomical observatory (Vernin and Muñoz–Tuñón, 1994; Klückers et al., 1998). These papers clearly show that the turbulence can occur at many strengths and heights in the atmosphere.

Variation of strength with height of large bulk motions, commonly called atmospheric turbulence can be experienced whilst flying. In a light aircraft during

a hot summers day, near the ground, *e.g.* on approach to land, the differential heating of areas of the ground causes turbulence which can be quite bumpy, uncomfortable and troublesome for the student pilot. However, at heights above a few thousand feet the air is less disturbed and the flight smoother. At the other extreme, passenger aircraft make use of high–speed jetstreams to cut journey times and reduce fuel consumption when cruising at altitudes of around 35 thousand feet. However, when climbing into these streams, quite severe turbulence can be encountered caused by the shear at the boundary of the stream and the lower, slower air.

Large scale bulk motion of air, as experienced by aircraft is not necessarily the cause of phase variation in the propagation of light. In this work, atmospheric turbulence refers to that turbulence which causes variation of the refractive index of air through changes of density caused by frictional heating of packets of air.

As the atmospheric turbulence introduces phase variations in an incident plane wave, the phase of all wavelengths of light is affected. The distortion introduces smaller relative effects for light of longer wavelengths, and results in the wavelength dependence in the definition of r_0 in equation 3.3.

r_0 is a useful quantity in determining the performance one can expect from a ground–based astronomical telescope. It can be thought of as the diameter on the sky over which the atmospheric phase distortion is essentially uniform and can be identified with the typical size of the "boiling" cells as described earlier.

r_0 is defined as that size over which the wavefront variance is 1 rad^2. If the wavefront variance is known, then the comparative performance with the diffraction limiting case, the Strehl ratio (section 3.3.2) is found through

$$S = e^{-\sigma^2} \ . \tag{3.4}$$

If the telescope diameter, D, is significantly less than r_0, then over the aperture, the phase of the wavefront is approximately constant. The images formed by such a telescope will be diffraction limited ($S \approx 1$). A telescope with $D = r_0$ will image with S = 0.37. A telescope with $D > r_0$ will encounter more than one "seeing cell", each of which will limit the resolution to the diameter of that cell. The image is built up from a superposition of images from each cell, with $S \ll 1$. The width of images can be predicted using (Beckers, 1993):

$$\Omega \approx \frac{\lambda}{r_0}, \qquad r_0 < D, \tag{3.5}$$

$$\Omega \approx \frac{\lambda}{D}, \qquad r_0 \geq D, \tag{3.6}$$

where Ω is the full width at half maximum (FWHM) of the image. The $r_0 < D$ case describes large astronomical telescopes. As $r_0 \propto \lambda^{\frac{6}{5}}$, this results in a wavelength dependence for image size (Woolf, 1982):

$$\Omega \propto \lambda^{-0.2}, \tag{3.7}$$

a relation which was confirmed by Boyd (Boyd, 1978).

Good seeing on Mauna Kea has $r_0(0.55\mu m) = 25cm$ (Racine et al., 1991). Observing at this wavelength with a zero–aberration telescope and imager gives an

image size of 0.45″, whilst observing in the near–infrared K–band ($2.2\mu m$) with the same perfect telescope would produce images of width 0.34″.

Increasing telescope diameter beyond r_0 does not increase the resolution available, as this is limited by the atmosphere. It does however drastically reduce integration times on objects. The $10m$ Keck telescopes have $\sim$ 1600 times the collecting area of a similar reflecting telescope of diameter $25cm$, decreasing integration time by this factor.

3.2.1 Measuring r_0

As r_0 is such a direct indicator of the resolution obtainable with a large telescope, quantitative measurements are now being required from telescope control systems. The Acquisition and Guidance Control Systems being developed for the Gemini telescopes will not only control the pointing of the telescope, but provide fast tip/tilt/focus measurements to the Telescope Control System (TCS), slower active Optics (aO) measurements to correct the figure of the primary, *and* provide an estimate of r_0 to the TCS.

The fast tip/tilt/focus (FTTF) corrections form a very low–order Adaptive Optics (AO) system, more of which anon. The camera being employed in the FTTF system makes many images of the guide star on a detector. Each image samples an area on the pupil of the telescope, therefore each image samples a different piece of the turbulence. Rapid readout of the detector will enable the motion of

each spot to be observed.

Estimates of r_0 can be made in the following manner. One can use the motion of a single image (Martin, 1987):

$$\sigma_{abs} = 0.0425 \left(\frac{\lambda}{0.5\mu}\right) \left(\frac{r_0}{1m}\right)^{\frac{5}{6}} \left(\frac{d}{1m}\right)^{\frac{1}{6}}, \tag{3.8}$$

where σ_{abs} is the mean square position of the image in arcseconds, and d is the diameter of the aperture, or the motion of pairs of spots in the direction of the sub aperture separation (Sarazin and Roddier, 1990):

$$\sigma_l^2 = 2\ \lambda^2\ r_0^{-\frac{5}{3}}\ [0.179\ d^{-\frac{1}{3}} - 0.0968\ r^{-\frac{1}{3}}], \tag{3.9}$$

where σ_l^2 is the variance of the differential image motion in arcseconds^2 between apertures of diameter d and separation r (a similar relation describes the transverse component). Sarazin and Roddier present measurements of image FWHM which indicate that r_0 varies rapidly: some excursions can be by as much as a factor of 2 in a matter of minutes.

The single aperture method of measuring r_0 is susceptible to telescope windshake. For this reason, the differential method is usually preferable. However, the differential method does not allow investigation of the image motion caused by the atmosphere which is possible with the single aperture.

The typical timescale over which the wavefront changes over the telescope pupil is determined by the speed at which the turbulence is moving across the field of view. As r_0 describes the size of a patch of turbulence over which there is

essentially no distortion, the timescale for variations in the wavefront is related to the time taken for one cell of size r_0 to be replaced by another (Beckers, 1993):

$$\tau_0 \approx 0.314 \left(\frac{r_0}{v_{wind}} \right). \tag{3.10}$$

Inputting sensible values for $r_0 = 25cm$ and $v_{wind} = 10ms^{-1}$ gives a timescale of $0.008s$. It is clear that the wavefront distortions can change very quickly.

3.2.2 Modal Representation of the Atmosphere

When describing the shape of a wavefront $W(\underline{r})$, it is convenient use a set of orthogonal modes $Z_i(\underline{r})$:

$$W(\underline{r}) = \sum_{i}^{\infty} a_i \, Z_i(\underline{r}), \tag{3.11}$$

where a_i are the coefficients of the modes. The most commonly used set of modes is the Zernike polynomials (Noll, 1976). For Kolmogorov turbulence, Noll quantifies the contribution of each mode to the aberration. In general terms, the higher the order of the mode (i), the smaller the contribution to the distortion.

3.3 Adaptive Optics

Adaptive Optics (AO) is the name given to systems designed to measure and compensate in *real time* for distortions introduced to the phase of light as it propagates through a medium. In the case of propagation through the atmosphere, the timescale for variation of the wavefront distortion is of order a few

milliseconds, requiring that an AO system update at a frequency of $\sim 1kHz$ so that the variation can be tracked.

3.3.1 The Perfect Adaptive Optical System

Hold the page at arms length, close one eye and look at this dot —

•

Bring the paper closer to you, and then take the paper away. You still see the dot and it is still in focus. By altering the distance of the paper from your eye, you are forcing your eye to alter its focal length. Your eye is constantly adapting to the prevailing conditions to give you the clearest vision possible. What is perhaps more remarkable is that on the whole, you do not have to worry about whether your eye is in focus or not. It just is, letting you get on with notching up your second century of the match.

In the same way, a perfect Adaptive Optics system for astronomical observing is one where astronomers use the telescope in exactly the same way that they would without Adaptive Optics. The only clue to its presence should be a measurable improvement in the quality of observations compared to telescopes without AO.

3.3.2 Quantifying Adaptive Optics Performance

The traditional measure of the resolution obtained with an astronomical telescope is the full width at half maximum (FWHM), Ω. As the quality of an atmospherically distorted image improves, a diffraction–limited core emerges in the centre of the image. Once formed, the FWHM of this core is essentially constant, and the FWHM measure is insensitive to further improvements in performance.

For this reason, the Strehl ratio, $\mathcal{S}$, is used instead. It is the central intensity of the image (Im) scaled by the central intensity one would expect given the diffraction–limited (DL) case (Roddier, 1993):

$$\mathcal{S} = \frac{I_{\mathrm{Im}}^{max}}{I_{\mathrm{DL}}^{max}}. \tag{3.12}$$

3.4 Components of an Adaptive Optics System

The original proponent of Adaptive Optics was Babcock (Babcock, 1953). He proposed a complete system, the structure of which is the basis for all astronomical AO systems currently in operation (Tyson, 1991; Alloin and Mariotti, 1993; Beckers, 1993; Ealey and Merkle, 1994; Cullum, 1995; Wizinowich et al., 1996b).

Figure 3.1 shows an outline of an AO system. Light from the telescope is reflected off a deformable mirror (DM), the figure of which is controllable. If the figure of the mirror is not flat, the shape of the reflected wavefront is different

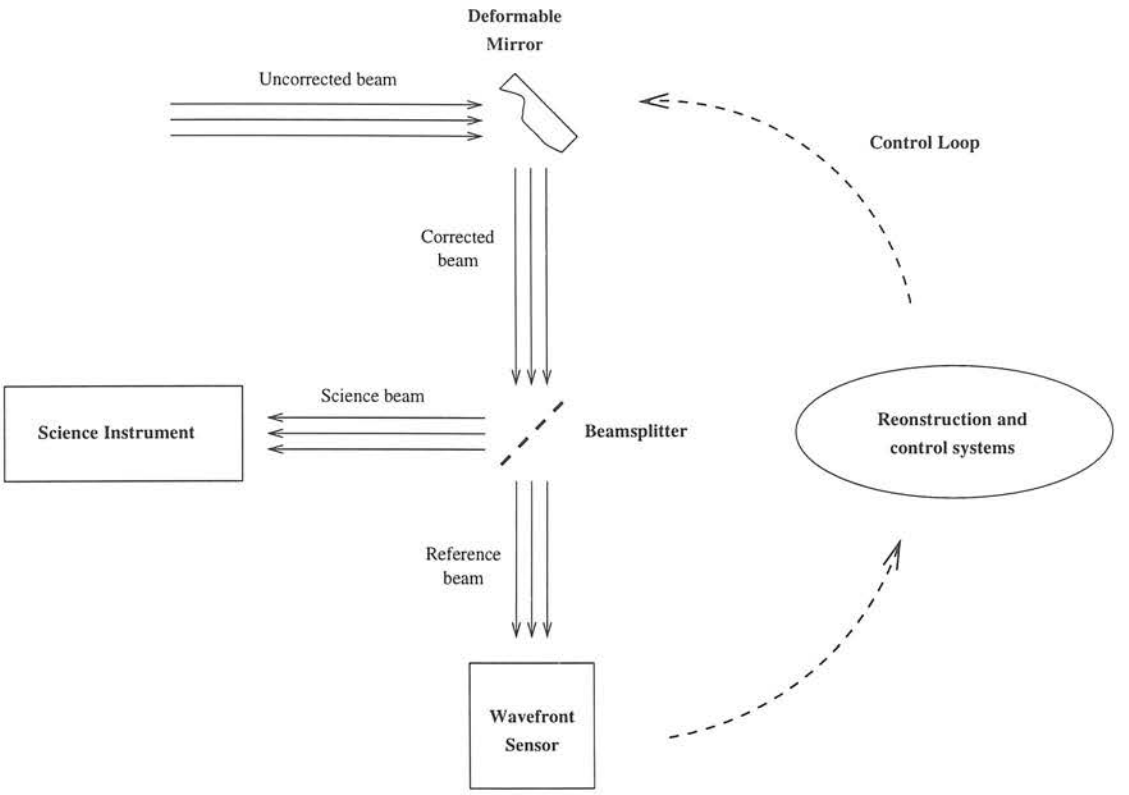


Figure 3.1: A schematic Adaptive Optics system. Light from the telescope is split between a science instrument and a wave–front sensor. The signals from the sensor are converted into actuator signals to drive the deformable mirror to correct the distorted wavefront. This flow of information forms a closed loop so that any error in correction can be determined and corrected.

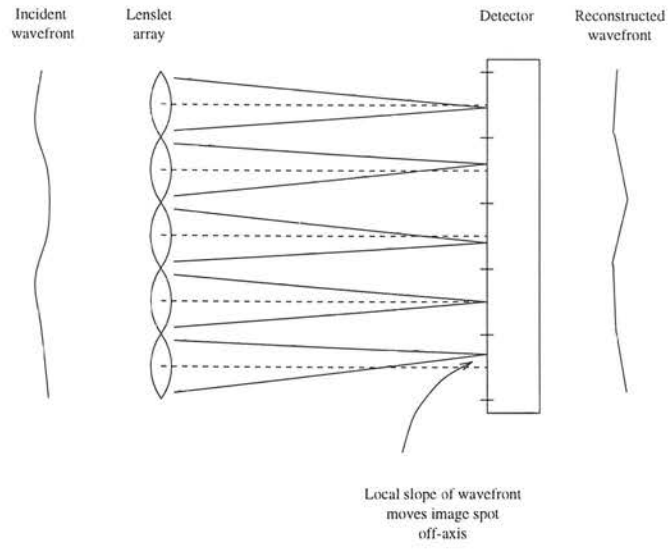


Figure 3.2: A Shack–Hartmann sensor.

to the incident wavefront. At the beamsplitter, part of the light (usually the infrared) is directed into the science instrument with the remaining visible light continuing to the wavefront sensor (WFS). The WFS measures the shape of the wavefront and hence the distortions that remain after correction by the DM. A non–flat wavefront causes correction signals to be sent to the DM through the control electronics. The system therefore forms a closed servo–loop that drives the corrected wavefront to be as flat as possible.

Astronomical Adaptive Optics systems come in two flavours — Shack–Hartmann (SH) sensor and Curvature Sensor (CS) systems. The fundamental differences between the two are the wavefront sensors and deformable mirrors.

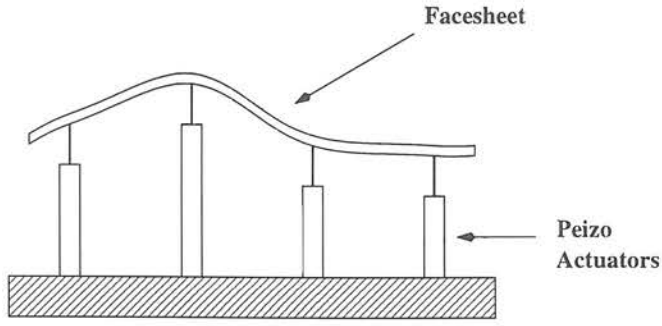


Figure 3.3: Cross–section through a stacked actuator mirror.

3.4.1 Shack–Hartmann AO System

The wavefront sensor in a Shack–Hartmann system places an array of lenslets at a plane conjugate to the telescope pupil in the optical path (Tyson, 1991). Each lenslet defines a subaperture, and this creates an array of images on the detector, usually a CCD with a few electrons read–noise — figure 3.2. An aberration in the incoming wavefront causes the spots to move off their local optical axis. The vector of displacement of each spot is directly related to the mean gradient of the phase of the wavefront over the corresponding lenslet subaperture. Thus a Shack–Hartmann sensor allows one to approximate the incoming wavefront as a series of inclined planes.

A Shack–Hartmann sensor is normally teamed with a continuous faceplate stacked actuator mirror — figure 3.3. A thin, flexible facesheet is mirror coated on one side. Attached to the other are a set of piezo–electric actuators which push the mirror. Varying the voltage on the actuators allows the shape of the mirrored surface to be changed. The mirror is a continuous surface (Ealey and Washeba,

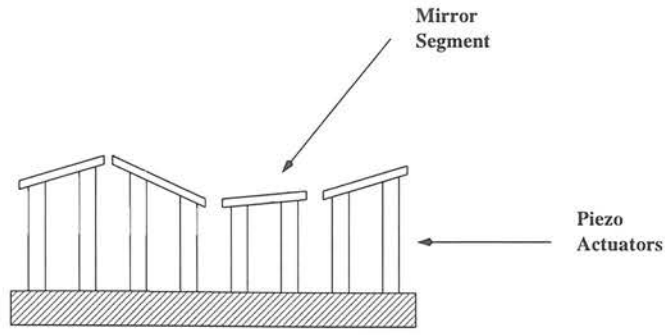


Figure 3.4: Cross–section through a segmented faceplate mirror.

1990).

Working SH systems using stacked actuator systems have been built (Fugate et al., 1991; Brase et al., 1994; Rousset et al., 1994).

Segmented faceplate mirrors are also used with Shack–Hartmann sensors. A segmented faceplate mirror is usually an array of small, rigid square mirrors, each of which can move independently in tip, tilt and piston fashion. A segmented mirror is not constrained to represent a continuous surface and as such can be fitted to an arbitrary surface with better success than a continuous surface mirror (Hurburd and Sandler, 1990).

Shack–Hartmann systems using segmented mirrors have been built and operated successfully: MARTINI (Doel et al., 1995) and ELECTRA (Buscher et al., 1995). The UK adaptive optics programme is building a similar facility system (NAOMI) for common use on the William Herschel Telescope based around the segmented ELECTRA mirror (Gilmore et al., 1995).

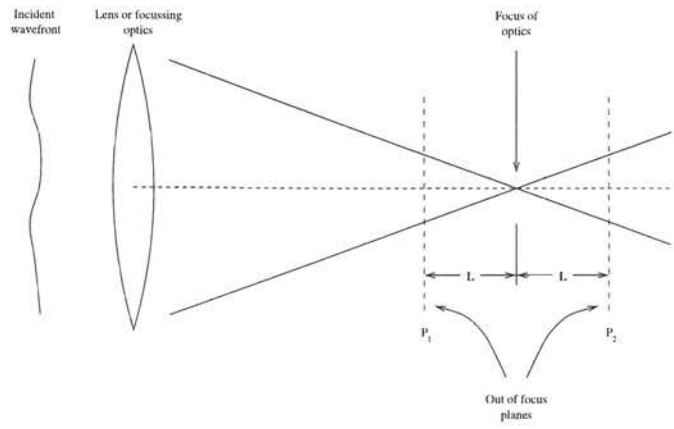


Figure 3.5: A Curvature Sensor.

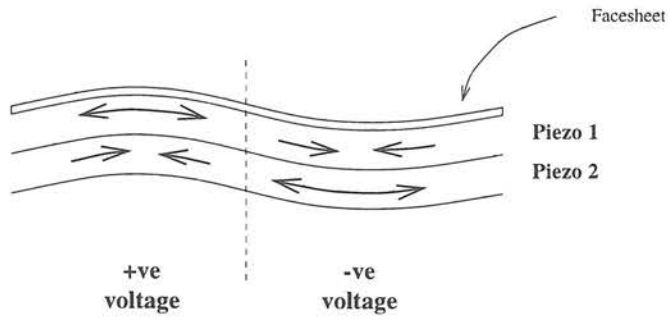


Figure 3.6: Cross–section through a bimorph mirror. Piezos 1 and 2 have opposite polarity so that when a voltage is applied one contracts whilst the other expands allowing a greater range of curvatures to be obtained.

3.4.2 Curvature Sensor AO System

The wavefront sensor in a Curvature Sensor system, as the name suggests, measures local wavefront curvature. This is in contrast to the local wavefront gradient as measured by a Shack–Hartmann sensor. Comparison of the out of focus images of a star, either side of the focal plane, allows measurement of the wavefront curvature (Roddier, 1988). The detector of choice for a Curvature Sensor is the avalanche photo–diode (APD) which has zero read noise.

Curvature Sensors are usually teamed with bimorph mirrors — figure 3.6. In these mirrors, two piezos of opposite type are sandwiched together and separated into elements to which voltages are applied. Each element behaves like a bi–metallic strip causing curvature in the mirrored faceplate. As with stacked actuator mirrors, bimorph mirrors are constrained to represent a continuous surface (Kokorowski, 1979; Steinhaus and Lipson, 1979).

Working systems using Curvature Sensors have been built (Roddier et al., 1991; Close et al., 1996; Rigaut et al., 1996).

3.4.3 Common AO system components

Several parts of an AO system are common to both Shack–Hartmann and Curvature Sensors, and will now be discussed.

Beamsplitter

The beamsplitter separates the light from the telescope into two beams. For optical wavelength adaptive optics, it could take the form of a half–silvered mirror. For astronomical applications, the beam splitter usually takes the form of a dichroic. The visible light is directed into the WFS whilst the infrared light continues to the science instrument.

The reasoning behind the wavelength nature of this split is twofold. First, visi-

ble CCD's and APD's have significantly less readnoise than comparable infrared detectors which results in better signal to noise from the WFS.

Secondly, r_0 is significantly larger at infrared wavelengths, *e.g.* $r_0(0.55\mu m) = 0.25cm$, but as $r_0 \propto \lambda^{\frac{6}{5}}$, $r_0(2.2\mu m) = 132cm$. This means that τ_0 (section 3.2.1) and θ_0 (section 3.5.4) are much larger, allowing AO systems to be built which run at a slower loop speed, with fewer subapertures, for the same performance benefit as in the optical.

Science Instrument

The science instrument should be designed with adaptive optics observations in mind. For example, the introduction of the tip/tilt secondary on UKIRT has improved the quality of images so that the existing infrared camera, IRCAM3 with a plate scale of 0.286″/pixel, is now regularly undersampling the point spread function. To remedy this, UFTI, with a plate scale of 0.090″/pixel, is about to be commissioned.

Reconstruction and Control Systems

The purpose of the electronics of the AO system is to convert the WFS measurements into control signals of the deformable mirror. Although straightforward in principle, there is much subtlety involved in the implementation of the control loop algorithms. The work presented in the remainder of this thesis is directed

at improving the performance of the control loop. Before the control loop is discussed in more detail in section 3.6, the limitations of adaptive optics will be outlined.

3.5 Limitations of AO Systems

As mentioned in section 3.3, for an adaptive optics system to accurately sample and correct for any changes in atmospheric distortions of starlight requires the system to operate on a similar timescale to that of the atmosphere. There are, however, a number of limiting factors which must be addressed when assessing the performance one can expect from an adaptive optics system.

3.5.1 Resolution of WFS

The wavefront sensor is the apparatus which measures the wavefront distortion. As such, the resolution to which the shape of the incoming wavefront can be measured is limited by the resolution of the WFS. In the Shack–Hartmann sensor case, an 8×8 subaperture arrangement will give twice the resolution in the x– and y–directions of a 4×4 arrangement.

With the WFS limiting the measurement resolution, the choice of deformable mirror employed with the system has to reflect the limit in resolution. In the SH case, a continuous faceplate mirror with 9×9 actuators arranged to corre-

spond with the corners of the subapertures would be the highest fidelity mirror worthwhile teaming with an 8×8 subaperture sensor.

3.5.2 Subaperture Illumination

In principle, the greater the number of subapertures, the better the performance available through greater sampling of the wavefront. However, moving from a 4×4 to an 8×8 subaperture arrangement on the same telescope necessarily requires that the illumination of each subaperture in the 8×8 case reduces to $\frac{1}{4}$ of that with the 4×4 sensor.

For the same guide star illumination, the signal to noise of the images in the subapertures may drop below that which is useful, actually leading to a penalty in performance.

3.5.3 Guide Star Magnitude

In the same manner as varying the number of subapertures, varying the intensity of the guide star has direct bearing on the performance of the AO system through the signal to noise in the subaperture images. There will be a guide star magnitude above which there is no benefit to be had from using an AO system directly with that object. There may however be a bright object nearby on the sky which would be able to guide the AO system whilst the science instrument concentrated on the fainter object.

3.5.4 Isoplanatism

The radius on the sky over which the atmospheric distortion can be thought of as constant is called the *isoplanatic patch* (Beckers, 1993). This is approximated by:

$$\theta_0 = 0.314 \frac{r_0}{H}, \tag{3.13}$$

where H is the effective distance of the seeing of scale r_0:

$$H = \sec\zeta \left(\frac{\int C_N^2 \, h^{\frac{5}{3}} \, dh}{\int C_N^2 \, dh} \right)^{\frac{3}{5}}, \tag{3.14}$$

where ζ is the angle of the source from the zenith.

For $H = 8km$ and $r_0(0.55\mu m) = 25cm$, this gives $\theta_0 \approx 4''$. The wavefront variance increases as the separation between the reference and guide sources due to the decorrelation of the atmosphere:

$$\sigma^2 = \left(\frac{\theta}{\theta_0} \right)^{\frac{5}{3}} \tag{3.15}$$

When $\theta = \theta_0$, for an AO system which can make perfect corrections at $\theta = 0$, there is the same performance decrease as an uncorrected telescope with $D = r_0$. As the separation of the sources increases on the sky, then the correlation of the wavefront between the science and guide sources breaks down further, and the performance of the imaging will become poorer.

3.5.5 WFS Integration Time

For a given subaperture arrangement and guide star magnitude, the wavefront sensor integration time also has a bearing on the performance available from an AO system. An integration time which is too short will not allow enough time for good signal to noise to be established in the subaperture images, whereas an integration time that is too long will not be able to freeze the effects of the atmosphere. There is a balance to be struck between a long integration time to get a good image and a short integration time to sample the turbulence with good temporal resolution.

A faint guide star will necessarily increase the optimum integration time. The distortions measured by the WFS will be an average of the distortions occurring over the integration time. For slowly varying distortions, this may still be a good approximation, but faster variations will change between integration periods, and the fastest variations will average out to zero as turbulence is a random phenomenon.

Thus, as the guide star becomes fainter and the integration time is increased to keep the performance optimal, there will be a drop in system performance through temporal averaging. The work presented here is aimed at alleviating this effect.

3.5.6 Numerical Models of AO systems

The issues described above make the choice of the parameters for a proposed AO system fairly daunting. A great deal of work has been done in building numerical models of AO systems to estimate the probable performance of a system (Roddier et al., 1991; Wampler et al., 1994; Gavel et al., 1996; Rigaut et al., 1997). For a good estimation of the likely performance from such a system, a good model of the atmosphere is required. In most cases, phase screens with the Kolmogorov structure function are used — equation 3.2 (Lane et al., 1992; Glindemann et al., 1993; Roggemann et al., 1995).

Work has also been done to quantify the theoretical degradations caused by angular isoplanatic and windspeed effects (Wilson and Jenkins, 1996). It would be safe to state that the general performance characteristics of AO systems and their limitations are well understood.

The numerical model of an adaptive optics system used in this work was very kindly supplied by François Rigaut (Rigaut et al., 1997) — see sections 3.6 and 3.11. The version supplied is written in the Interactive Data Language (IDL) (RSI, 1994). The phase screens supplied with the model show good agreement with the Kolmogorov $r^{\frac{5}{3}}$ power law. The model is a powerful investigative tool when run in batch mode, but also has an elegant graphical user interface for easy interactive investigation.

3.5.7 Sky Coverage with AO Systems

This understanding of the limits of AO system performance, coupled with knowledge of the distribution and magnitudes of stars, allows an investigation into the likely amount of sky which can be used for AO compensated astronomical observations (Rigaut, 1993; Ellerbroek et al., 1995). Ellerbroek has shown that if one wishes to obtain a Strehl of 0.3 in the H near infrared band, there is about an 8% chance of finding a suitable reference star with which to guide an AO system. In the K band, this increases to $\sim$ 30%. Especially in H, this is a small fraction of the sky. Given that there are many more faint stars than bright stars, any improvement in the performance of the AO system would be welcome as it would increase sky coverage.

The paltry nature of the sky coverage available with natural guide stars has led to the proposing of the use of laser beacons to provide reference sources for AO (Freedman, 1993; Greenwood and Parenti, 1993). The work described by Friedman was in connection with the construction of a system for the Lick Observatory (Oliver et al., 1996). Work is underway to provide a similar facility for the Keck telescopes (Wizinowich et al., 1996a).

The laser excites the layer of sodium atoms in the atmosphere at a height of around $80km$. These then radiate isotropically with some being collected by the astronomical telescope. Placing the laser beacon close to the science object enables good atmospheric compensation.

Another working system exists at the Starfire Optical Range near Albuquerque, New Mexico, where the US Air Force image objects in low Earth orbit (Fugate et al., 1991). This presents a difficult problem as the objects of interest can move rapidly across the sky making the use of stars as beacons unfeasible.

3.6 Control Loop Algorithm

The control loop algorithm of Rigaut's model will now be discussed.

3.6.1 Generation of the Interaction Matrix

When the model AO system is running, we require the WFS signals to be converted into voltages to apply to the actuators of the DM as described by:

$$\underline{v} = \mathcal{C} \cdot \underline{s} \tag{3.16}$$

where $\underline{v}$ are the voltages to apply to the mirror, $\underline{s}$ are the WFS sensor signals and $\mathcal{C}$ is the *Control Matrix* which converts sensor signals into actuator voltages.

This cannot happen unless there is knowledge of the response of the WFS to known inputs on the DM. Note that "DM" in this context refers to both the actual DM and, if appropriate, a separate tip/tilt mirror to provide enough stroke to correct for image motion.

The response of the WFS, $(d_1, d_2, .. d_N)$, to each DM actuator is determined and the WFS output is recorded as a column in the *Interaction Matrix* $\mathcal{D}$. The inter-

action matrix has M columns, one for each actuator in the mirror and N rows, the length of the column matrix outputted by the WFS as shown in equation 3.17.

$$\begin{pmatrix} s_1 \\ s_2 \\ . \\ . \\ s_N \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} & . & . & d_{1M} \\ d_{21} & d_{22} & . & . & d_{2M} \\ . & & & & . \\ . & & & & . \\ d_{N1} & d_{N2} & . & . & d_{NM} \end{pmatrix} . \begin{pmatrix} v_1 \\ v_2 \\ . \\ . \\ v_N \end{pmatrix} \tag{3.17}$$

or in shorthand:

$$\underline{s} = \mathcal{D} . \underline{v} \tag{3.18}$$

If only one actuator is pressed, *i.e.* one element of $\underline{v}$ is non-zero, then we get the response of the WFS to that actuator. If more than one is pressed, each element in $\underline{s}$ is a linear combination of its response to each press on each actuator. Thus the behaviour of the WFS in response to the DM is characterised by the interaction matrix.

It should be noted that the determination of the response of the system to actuator inputs is performed in an idealised environment. A blurred source is modelled to give an image of approximately the same size as one in good seeing in each subaperture of the WFS. A high light level is used, giving very good signal to noise in each subaperture image.

3.6.2 Calculation of the Control Matrix

As discussed in section 3.6.1, when the AO system is in operation, we require to solve

$$\underline{v} = \mathcal{C} . \underline{s} \tag{3.19}$$

It is clear that the control matrix is the inverse of the interaction matrix, *i.e.*:

$$\mathcal{C} \equiv \mathcal{D}^{-1}. \tag{3.20}$$

Inversion of matrix $\mathcal{D}$ is performed in the model using singular value decomposition (SVD) (Press et al., 1988). SVD is a reliable way of determining the inverse of a matrix, or for a matrix which is close to singular, a pseudo–inverse. For matrix $\mathcal{D}$, SVD finds $\mathcal{V}$, $\mathcal{U}$ and $\underline{w}$ such that:

$$\mathcal{D}^{-1} = \mathcal{V} \,.\, [\mathrm{diag}(1/w_j)] \,.\, \mathcal{U}^T. \tag{3.21}$$

$\mathcal{U}$ and $\mathcal{V}$ are sets of orthonormal vectors and $\underline{w}$ is a set of eigen–values. The columns of $\mathcal{U}$ whose corresponding w_j's are non–zero are an orthonormal set of basis vectors which span the range of $\mathcal{D}$, and the columns of $\mathcal{V}$ whose corresponding w_j's are zero are an orthonormal basis for the nullspace of $\mathcal{D}$.

In the case of an AO system, there are certain modes of the mirror to which the WFS has no response, or though real modes, give the same response as other modes. An example of a mode which has no response is one of the waffle modes which has opposite voltages applied to adjacent actuators leaving zero mean slope on the mirror in each sub–aperture. Tip plus a waffle mode, for a similar reason, will have the same response as tip. It is reassuring that SVD assigns these modes the smallest eigen–values.

When constructing $\mathcal{D}^{-1}$ by equation 3.21, modes of the nullspace have their $(1/w_j)$ set to zero lest infinities creep in. In the same way, the degree of correction

of the AO system can be reduced from that which is theoretically possible given the number of actuators in the mirror by setting more of the $(1/w_j)$'s to zero. It is usual to suppress those modes with the smallest w_j's as they have a lesser role to play in the solution. Varying the degree of suppression allows one to tune the degree of correction to allow for anisoplanatic and other effects — section 3.7.

For a model system with a realistic number of actuators in the DM, the generation of the control matrix is a computationally intensive task. For this reason, the control matrix for each WFS/DM configuration can be saved to file for later use.

3.6.3 Operation of the Control Loop

When modelling the operation of the AO system the simulation executes the following loop. First, move the phase screen(s) the appropriate amount and determine their combined effect across the telescope pupil to model the atmosphere $A(\underline{r})$. The DM is conjugate with the telescope pupil, so its shape $DM(\underline{r})$ can be added to the atmospheric phase to give the correction error signal $E(\underline{r})$:

$$E(\underline{r}) = A(\underline{r}) + DM(\underline{r}) \tag{3.22}$$

The lenslets of the WFS are conjugate with the telescope pupil, and the correction error signal is converted into WFS image spots using Fourier propagation at the sensing wavelength. The WFS detector is modelled to have a realistic amount of read noise. Meanwhile, the science image is formed at the appropriate wavelength. The spot positions, measured with respect to their local optical axes, $\underline{\delta s}$, represent

the error in getting the wavefront flat, and should be small if the system is performing well. The positions are determined by the quad-cell algorithm (Tyson, 1991) and are presented to the control loop. The amount of DM correction to apply is determined by

$$\underline{\delta v} = \mathcal{C}\,\underline{\delta s} \tag{3.23}$$

and the mirror shape is updated by

$$\underline{v}_{new} = \underline{v}_{old} + g \cdot \underline{\delta v} \tag{3.24}$$

where g is the loop gain. Thus the loop is complete. Any noise in $\underline{\delta s}$ will propagate through this process, resulting in an inaccurate estimate for $\underline{v}_{new}$. The greater the noise, the worse this effect.

At each iteration, the Strehl of the instantaneous science image is calculated. The instantaneous science images are coadded to build up a long exposure image, the Strehl of which provides our measure of the AO system's performance over the run.

3.7 Optimisation of AO Performance

This subject has already been touched upon in section 3.5. The goal is to find the parameters of a given AO system which will provide the best performance for a specific set of observing conditions.

3.7.1 Integration Time

There is a balance to be struck between increasing integration time to overcome measurement noise and deceasing integration time to freeze the variations in atmospheric aberrations.

3.7.2 Modal Optimisation

The control loop in Rigaut's model solves directly for the sensor voltages. In the process, the modes of the system are determined (columns of $\mathcal{U}$). Some systems however employ a two–stage process where the shape of the distortion measured by the wavefront sensor is determined. The usual method is to obtain the coefficients for a set of orthogonal modes such as the Zernike polynomials (Noll, 1976) through a matrix operation as described above. The voltages to apply to the mirror can then be obtained through a second matrix operation. In either case, the number of modes of correction can be altered although it is achieved more easily in the latter case.

Coupled with the integration time issue is the question of the degree of correction performed by the AO system. The lowest order modes have the greatest power (Noll, 1976) and the longest timescale. In section 3.6.2, it was noted that if the integration time is large, then there are atmospheric effects which are not measured due to their timescale being smaller than that measurable. These effects which are not measurable correspond to the higher order modes. As there is no

signal with which to measure these modes, attempting to determine them can only introduce noise in to the estimate for the DM voltages. Optimisation of the gain of each of individual modes, with the gain of the highest modes being set at zero, can therefore lead to improvements in system performance (Gendron and Léna, 1994).

In a similar manner to the temporal decorrelation of the atmospheric modes, increasing the separation of the guide and science sources on the sky will lead to spatial decorrelation. Again, the suppression of higher order modes can yield a boost in system performance (Wilson and Jenkins, 1996).

3.8 Atmospheric Prediction

Given an AO system working with a reasonably optimal set of parameters, can a further improvement be made in the performance the AO system?

One line of work is to investigate the possibility of temporal prediction of the atmosphere to overcome the delay between the measurement and correction of atmospheric distortion. This may be a fruitful venture as the atmosphere turbulence, as measured by $C_N^2(h)$ occurs in a few distinct layers. These layers can be thought of as containing essentially fixed turbulence which is blown across the telescope pupil a wind. Making a measurement of the turbulence in one subaperture, together with knowledge of the velocity of the wind, allows a prediction of

the phase to be made in a second downwind subaperture.

Experiments on measurements of wavefront tilt data taken with the Multiple Mirror Telescope (MMT), Mt. Hopkins, Arizona in *open loop* (*i.e.* no adaptive optics correction) have shown that the tilts are predictable using simple neural networks of linear predictors (Jorgenson and Aitken, 1992). Using data from the COME–ON system at the 3.6m ESO telescope at La Silla, again in open loop, Aitken and McGaughey have confirmed these findings and conclude that short–term prediction could overcome the inherent time delays in an AO system (Aitken and McGaughey, 1995). Their results suggest that in good seeing, neural network and linear predictors are comparable, whereas for poor seeing, neural networks outperform linear predictors by a factor of 2 in reducing the root mean square error signal.

Aitken and McGaughey have gone on to characterise the atmospheric statistics as being well described by fractional Brownian motion. They find that, in principle, the atmospheric distortions are not predictable, and that the predictability comes from the spatial filtering of the distortions by the Shack–Hartmann subaperture pupils (McGaughey and Aitken, 1997). It is odd to think that the process of performing a measurement is the source of the predictability in the process of interest.

3.8.1 Practical Implementation

The question investigated in this thesis is the following. Given that the measurements of atmospheric turbulence are predictable, can this knowledge be included in a practical (*closed loop*) implementation of an AO system to boost the performance? If so, quantify the potential benefits.

One proposition is to optimise the control matrix of the loop (equation 3.16) given prior knowledge of the statistics of the atmospheric effects. This has been developed for the Kolmogorov case (Wild, 1996). The question arises of what to do when the statistics are non–Kolmogorov. A more general approach would be more beneficial.

The work outlined above by Aitken and others shows that the atmosphere is predictable through measuring local wavefront slopes in open loop. In the case of an AO system working in closed loop, however, the measurements made by the the WFS are of the error signal: the difference between the shape of the DM and the incoming atmospheric distortion — equation 3.22.

The wavefront error signals may not be predictable for the following reason. Assume that the position adopted by the DM is the result of a prediction made by some predictive algorithm, and the next measurement of the wavefront error position is then performed. The error signal is therefore a difference between the atmosphere and the prediction of the predictor. For the next prediction to be

accurate, the predictor has to understand how both it and the atmosphere are behaving. Whether this is predictable or not will depend on the type of predictor, and is not easily known. What is certain however, is that the predictor is not concentrating specifically on the process which is known to be predictable — the measurements of atmospheric distortions.

How can the atmospheric distortions be recovered from the wavefront error signals?

Equation 3.22 can be re–arranged to give

$$A(\underline{r}) = E(\underline{r}) - DM(\underline{r}). \tag{3.25}$$

The behaviour of the atmosphere can therefore be recovered if the position of the DM is known accurately. In the case of piezoelectric stacked actuator mirrors, accurate positional estimates cannot be guaranteed as hysteretic effects can cause a discrepancy between the expected shape given the applied voltages and that actually attained by the mirror.

The ELECTRA mirror, to be used in NAOMI in the UK AO programme, has strain gauges on each segment of the mirror creating a positional servo–loop for each segment to ensure that the overall mirror figure is that which is desired. This results in a mirror with zero hysteresis, and allows equation 3.25 to be used to accurately reconstruct the behaviour of the atmosphere. The open loop case has been recovered from the closed loop measurements.

So, the answer to the initial question at the beginning of this section is the following. It should be possible to implement a predictive scheme in the NAOMI system for the WHT. This scheme is described in detail in section 3.10. Chapter 4 presents the results of a series of experiments which determine the type and configuration of the predictive algorithms and quantify the available performance improvement.

A second way of recovering the atmosphere from wavefront sensor error measurements is to use knowledge of the workings of the AO loop. This has been investigated in recently published concurrent work by a group at ONERA (Dessenne et al., 1997; Dessenne et al., 1998). Their method is closely related to the optimisation of the gain of individual modes of correction (Gendron and Léna, 1994). Through determination of the transfer function of an AO system, the closed loop WFS measurements can be converted into what would have been obtained in the open loop case.

3.9 Predictors

The two types of predictor investigated in this thesis will now be discussed.

3.9.1 Linear predictors

Linear predictors are algebraically simple beasts. A linear combination of previous measurements estimates the next:

$$y_n = \sum_{j=1}^{N} d_j \, y_{n-j} + x_n, \tag{3.26}$$

where y_{n-j} are known measurements, y_n is the next measurement, d_j are the linear prediction coefficients and x_n is the error in the prediction. Clearly, if $|x_n|$ is small compared with $|y_n|$, then the prediction is good. Ideally, one requires

$$|x_n| \ll |y_n| \tag{3.27}$$

for all n. The coefficients d_j are found through a maximum entropy method which characterises the signal in terms of a finite number of coefficients (Press et al., 1988). Linear prediction is not the same as linear extrapolation. The "linear" refers to the way in which the coefficients are combined — linear predictors can characterise very complicated functions.

The linear predictors described here are a limited set of the more general linear controller, of form

$$\mathcal{C}(z) = \frac{\sum_{i=0}^{q-1} a_i z^{-i}}{1 + \sum_{i=0}^{p-1} b_i z^{-i}}, \tag{3.28}$$

with $b_i = 0$. This simplified case was investigated as routines for implementing this predictor are readily available.

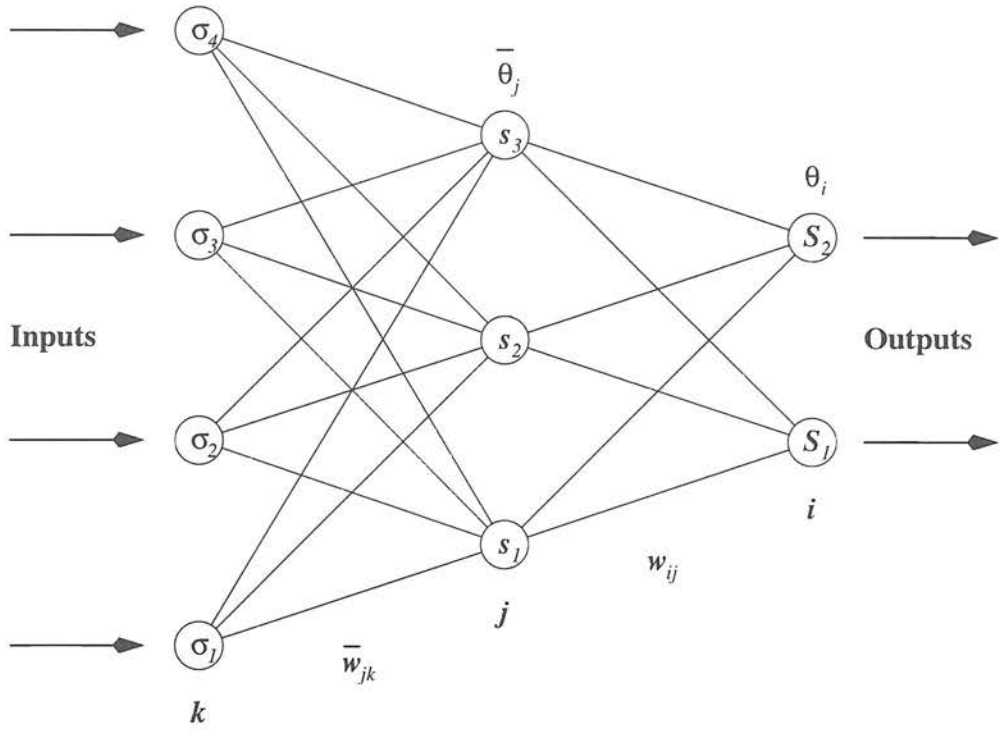


Figure 3.7: A simple *feed forward* neural network.

3.9.2 Neural Networks

Figure 3.7 shows a simple neural network with interconnected nodes. A datavalue (σ_k) is presented to each input and broadcast by each input node to the nodes in the second layer. Along the way, the datavalue is amplified or attenuated by a *weight* ($\overline{w}_{jk}$) which is intrinsic to each interconnection.

The hidden layer nodes total up the datavalues coming off the connections from the input layer and apply their own intrinsic *trigger levels* ($\overline{\theta}_j$) which give the total ($\overline{h}_j$). A non–linear function (*e.g.* tanh) is applied to this total to give the result (s_j). This is broadcast to the output nodes with weights (w_{ij}) applied along each connection. The output nodes perform the same procedure as the hidden

nodes.

This is notated:

$$S_i = f(h_i), \qquad h_i = \sum_j w_{ij} s_j - \theta_i, \tag{3.29}$$

$$s_j = f(\overline{h}_j), \qquad \overline{h}_j = \sum_k \overline{w}_{jk} \sigma_k - \overline{\theta}_j, \tag{3.30}$$

where $f(x) = \tanh(\beta x)$.

3.9.3 Training of Neural Networks

Appropriate choices of the weights and trigger levels enable the network to represent accurately the process detained in the training set. For this to be the case it is required that:

$$D\left[w_{ij}, \theta_i, \overline{w}_{jk}, \overline{\theta}_j\right] = \frac{1}{2} \sum_\mu \sum_i \left[\zeta_i^\mu - f(h_i^\mu)\right]^2, \tag{3.31}$$

is a minimum, where D is called the cost function and ζ_i^μ is the required value of the i^{th} output node in the μ^{th} case in the training set. The weights and triggers are variables which determine D so altering these will enable the minimum to be found.

3.9.4 Error Back Propagation

Error back propagation to train a network uses iterative gradient descent of the many dimensional surface D (Müller and Reinhardt, 1991). Considering the triggers in, and weights on the connections to, the output layer, the following is

obtained:

$$\begin{aligned}
\delta w_{ij} &= -\epsilon \frac{\partial D}{\partial w_{ij}} \\
&= \epsilon \sum_{\mu} [\zeta_i^{\mu} - f(h_i^{\mu})] f'(h_i^{\mu}) \frac{\partial h_i^{\mu}}{\partial w_{ij}} \\
&= \epsilon \sum_{\mu} \Delta_i^{\mu} s_j^{\mu}, \qquad (3.32)
\end{aligned}$$

$$\begin{aligned}
\delta \theta_i &= -\epsilon \frac{\partial D}{\partial \theta_i} \\
&= \epsilon \sum_{\mu} [\zeta_i^{\mu} - f(h_i^{\mu})] f'(h_i^{\mu}) \frac{\partial h_i^{\mu}}{\partial \theta_i} \\
&= -\epsilon \sum_{\mu} \Delta_i^{\mu}, \qquad (3.33)
\end{aligned}$$

where:

$$\Delta_i^{\mu} = [\zeta_i^{\mu} - f(h_i^{\mu})] f'(h_i^{\mu}), \qquad (3.34)$$

and ϵ is a gain factor and:

$$f'(x) = \beta(1 - f^2(x)). \qquad (3.35)$$

Considering the triggers in, and weights to the hidden layer in the same fashion:

$$\begin{aligned}
\delta \overline{w}_{jk} &= -\epsilon \frac{\partial D}{\partial \overline{w}_{jk}} \\
&= \epsilon \sum_{\mu,i} [\zeta_i^{\mu} - f(h_i^{\mu})] f'(h_i^{\mu}) \frac{\partial h_j^{\mu}}{\partial s_j} \frac{\partial s_j}{\partial \overline{w}_{jk}} \\
&= \epsilon \sum_{\mu,i} \Delta_i^{\mu} w_{ij} f'(\overline{h}_j^{\mu}) \frac{\partial \overline{h}_j}{\partial \overline{w}_{jk}} \\
&= \epsilon \sum_{\mu} \overline{\Delta}_j^{\mu} \sigma_k^{\mu}, \qquad (3.36)
\end{aligned}$$

$$\begin{aligned}
\delta \overline{\theta}_j &= -\epsilon \frac{\partial D}{\partial \overline{\theta}_j} \\
&= \epsilon \sum_{\mu,i} [\zeta_i^{\mu} - f(h_i^{\mu})] f'(h_i^{\mu}) \frac{\partial h_j^{\mu}}{\partial s_j} \frac{\partial s_j}{\partial \overline{\theta}_j}
\end{aligned}$$

$$= \epsilon \sum_{\mu,i} \Delta_i^\mu w_{ij} f'(\overline{h}_j^\mu) \frac{\partial \overline{h}_j}{\partial \overline{\theta}_j}$$

$$= -\epsilon \sum_\mu \overline{\Delta}_j^\mu, \tag{3.37}$$

with:

$$\overline{\Delta}_j^\mu = \left(\sum_i \Delta_i^\mu w_{ij} \right) f'(\overline{h}_j^\mu). \tag{3.38}$$

This method is recursive, enabling generalisation to many layers. The equations governing the state of a two hidden layer network are:

$$S_i = f(h_i), \qquad h_i = \sum_j w_{ij} s_j - \theta_i, \tag{3.39}$$

$$s_j = f(\overline{h}_j), \qquad \overline{h}_j = \sum_k \overline{w}_{jk} \overline{s}_k - \overline{\theta}_j, \tag{3.40}$$

$$\overline{s}_k = f(\overline{\overline{h}}_k), \qquad \overline{\overline{h}}_k = \sum_l \overline{\overline{w}}_{kl} \sigma_l - \overline{\overline{\theta}}_k, \tag{3.41}$$

with the adjustments defined as:

$$\delta w_{ij} = \epsilon \sum_\mu \Delta_i^\mu s_j^\mu, \qquad \delta\theta_i = -\epsilon \sum_\mu \Delta_i^\mu, \qquad \Delta_i^\mu = [\zeta_i^\mu - S_i^\mu] f'(h_i^\mu), \tag{3.42}$$

$$\delta \overline{w}_{jk} = \epsilon \sum_\mu \overline{\Delta}_j^\mu s_k^\mu, \qquad \delta\overline{\theta}_j = -\epsilon \sum_\mu \overline{\Delta}_j^\mu, \qquad \overline{\Delta}_j^\mu = f'(\overline{h}_j^\mu) \sum_i \Delta_i^\mu w_{ij}, \tag{3.43}$$

$$\delta \overline{\overline{w}}_{kl} = \epsilon \sum_\mu \overline{\overline{\Delta}}_k^\mu \sigma_l^\mu, \qquad \delta\overline{\overline{\theta}}_k = -\epsilon \sum_\mu \overline{\overline{\Delta}}_k^\mu, \qquad \overline{\overline{\Delta}}_k^\mu = f'(\overline{\overline{h}}_k^\mu) \sum_j \overline{\Delta}_j^\mu \overline{w}_{jk}, \tag{3.44}$$

where the additional bar in $\overline{\overline{w}}_{kl}$ denotes weights one layer further from the output layer than those in $\overline{w}_{jk}$ in a network with l inputs.

The adjustments to the weights and triggers depend on the error between the value outputted by the network and those required by the training set. The order in which the adjustments are calculated start with the output layer, with subsequent corrections filtering backwards through the layers and depending on

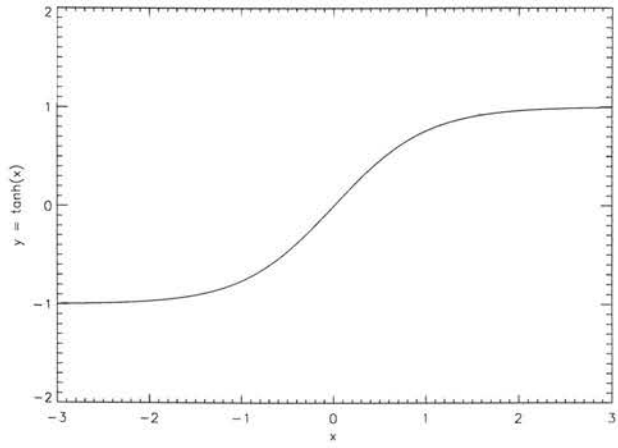


Figure 3.8: Plot of $y = \tanh(\beta x)$ with gradient through the origin, $\beta = 1$.

the adjustments in layers nearer to the output nodes. There is a wave of correction which travels from the output to the input nodes, the opposite direction to the data in normal operation of the network, hence the name *error back-propagation.*

3.9.5 Training Tricks

Scaling of inputs and outputs

The function applied in the hidden layer, tanh, is shown in figure 3.8. The range of x over which there is a response in y is the roughly the range $[-1, +1]$. Therefore it must be ensured that the totals h will usually be in this range. This is done by scaling the values in the training set so that they lie in the same range. This requires that there is knowledge of the extreme values of the process being characterised to scale the training set appropriately. In the time series prediction case, a period of time is set aside for determining the extreme values before the training set is recorded.

Initial network configuration

The initial values of the weights and triggers in a new network are usually assigned randomly. It is sensible to ensure that these numbers are small, so that large, time consuming corrections from a positive to a negative value are minimised as much as possible. The approach employed here was to use initial values sampled randomly from the range $[-0.1, +0.1]$.

Accelerated Back–Propagation

The calculated adjustments to the network parameters can be very large if the network is in a steep part of the surface D. Large adjustments can make the gradient descent unstable and prevent convergence into the required minimum. For this reason the gain ϵ is introduced. Only a fraction (usually $\sim$ 0.001) of the calculated adjustment is used to prevent the process running away.

Unfortunately, the speed of convergence of the training can be painfully slow if only this very small correction is made each training epoch. To speed things up the gain can be varied. The method employed here is to multiply the gain by a factor of 1.1 if the value of the function D was reduced by the corrections applied last training epoch. Thus, if the parameters are changing in the correct direction, the speed with which they vary is increased each epoch. However, if the value of D increased, then α is reset to its original value and the process started again. This method is called *accelerated back–propagation* and has a dramatic effect in

improving the speed of convergence.

Momentum

It is conceivable that the adjustments to the network parameters will cause the network to be in a local and not the global minimum of the function D. Adding a proportion of the previous correction gives the training process *momentum* which will help to kick the network through a local minimum, *i.e.*:

$$\delta w_{ij}^{(n)} = \epsilon \sum_{\mu} \Delta_i^{\mu} s_j^{\mu} + \alpha \delta w_{ij}^{(n-1)}. \tag{3.45}$$

Varying α varies the amount of momentum that is carried by the training process - too little and the network can still be trapped in a local minimum, too much and the network oscillates about the global minimum. A value of 0.6 was found to improve the training characteristics significantly.

Vector Prediction

The time series problem explored in this thesis requires that the network only predict one AO system iteration ahead (the scalar problem). It has been shown that training the network to predict two or more AO system iterations ahead (the vector problem) smoothes the function D and allows quicker convergence of the network (Gallant and Aitken, 1996). This was found to be the case though the effect is nowhere as marked as introducing accelerated back–propagation. All network training uses two output nodes, though when a prediction is made, only

the node corresponding to the next iteration is considered.

3.9.6 Other Training Techniques

A method called *simulated annealing* is another way to obtain convergence of a network (Hertz et al., 1991). A weight or trigger is selected at random and varied by a random, small amount. If there is a change for the better in terms of the function D, then the change is kept and the process is repeated. If the change is for the worse, then the change is accepted with a certain probability.

Many such incremental changes are made to the parameters, and over time the probability of accepting a modification which makes things worse is reduced. The role of accepting changes which make things worse is to allow the network to escape from local minima. Ultimately, only changes which make a benefit are chosen by which time the network should have converged. Simulated annealing is far more time consuming than back–propagation and though it was implemented and tested, was not considered further for this reason.

3.9.7 Time Series Prediction

The definition of the linear predictor explicitly states the time series prediction problem. The definition of the neural network however is more general. By the correct choice of the training set for the neural network, it is possible to train a neural network to solve the time series prediction problem.

The predictor is not explicitly "aware" that the problem it is solving contains a time series. The selection of training set data which forms a time series enables the predictor to determine the characteristics of the series and hence make a sensible prediction.

3.9.8 General Considerations for Predictors

The choice of the number of free parameters within the predictor algorithm has a direct bearing on the performance of the predictor. Varying the number of inputs and outputs affects not only the time taken to train the predictor, but depending on the characteristics of the process which one is trying to predict, it could influence the accuracy of the prediction.

If one selects too small a number of inputs, then the predictor may not be able to discriminate between two or more similar cases as it does not have enough information. Conversely, if one selects too great a number of inputs then there may be redundancy and computing time is wasted as some inputs offer no new information.

Similarly, the number of examples which are contained in the training set will also have influence on the performance of the prediction. If the set is too short, the predictor will not have access to the full range of behaviour of the process it is trying to learn and the characterisation will not be complete. Conversely, if the learning set is too long, then the time taken to train the predictor can become

unfeasible.

Clearly, there is a huge parameter space which one can explore in the quest to obtain the optimal parameters for a specific predictor for a specific problem. One could do a systematic search of the space but for complicated problems this becomes impractical.

A solution commonly employed is to find the best performers from a wide but sparse sampling of the parameter space. The parameters of the best performers are modified in some way, perhaps by swapping various parameters (cross–breeding) among the best performing cases and perhaps tweaking their values a little (mutation), and the problem is rerun with the new parameters. The process is repeated until a stable set of parameters is found. Such a scheme is called a *genetic algorithm* in that the rules for the swapping and changing of the parameters are thought to mimic the way DNA is modified through generations of species.

The major disadvantage of such an approach is that it still requires a huge amount of computing power as many generations of sets of parameters have to be created before a stable set is found. The work presented in Chapter 4 is restricted to a very limited set of parameters which have been found to give good results. It is perfectly possible that a large–scale search of the parameter space using genetic techniques would yield a more optimum configuration, but this was impractical given time and computing limitations.

3.9.9 Neural Networks in Astronomy

Using neural networks in astronomical applications is not restricted to prediction problems. Their use has been reported in determining the tip/tilt/piston corrections to necessary to co-phase the mirrors of the MMT (Angel et al., 1990). Other papers report determination of r_0 from Shack–Hartmann WFS spot positions (Montera et al., 1996a) and make centroid estimates for Shack–Hartmann WFS subapertures (Montera et al., 1996b).

3.10 Prediction Scheme

Previous work on prediction has focussed on individual wavefront slope measurements from the WFS (Jorgenson and Aitken, 1992; Aitken and McGaughey, 1995). The control loop of an AO system converts these slope measurements into incremental voltage corrections to change the shape of the deformable mirror. In the case of the NAOMI system, the shape of the mirror after the voltage corrections have been applied will accurately represent the system's approximation of the atmospheric distortion, given the most recent WFS measurement. It is the application of this most recent measurement to correct the atmosphere in the next measurement period which leads to the temporal degradation of the performance that prediction will attempt to compensate.

The aim is to make the addition of a predictor to the NAOMI system as straight-

forward as possible. It is envisaged as a "bolt on" addition to the existing control loop with as little interference as possible in the operation of the loop.

The information that is recorded for use in the predictor is the incrementally corrected mirror shape determined for use during the next measurement period. As this shape is the system's representation of the atmospheric distortion, it should be predictable. The shape of the surface can be decomposed into a sum of orthogonal modes; the choice here is to use the set of mirror modes. The coefficients of these modes are recorded for us in the predictor. If there is suppression of the higher order modes in the AO system, then the corresponding higher order mirror modes will be absent from the data presented to the predictor.

A novel feature about this implementation scheme is that the training set used to train the predictor is continually updated with the latest complete example of the behaviour of the atmosphere. The length of the training set is fixed to keep the computation required in each AO system iteration constant. The oldest data example is discarded to make way for the newest, and the predictor is re–trained at each AO system iteration so that the most recent behaviour of the atmosphere is characterised. The latest measurement of the atmosphere can the be presented to the predictor and a prediction obtained for the next AO system iteration.

A useful measure of how well the predictor is trained is to determine how well the predictor reproduces the training set and compare it to the performance of the *best is last*, standard AO system over the training set. If there is significant

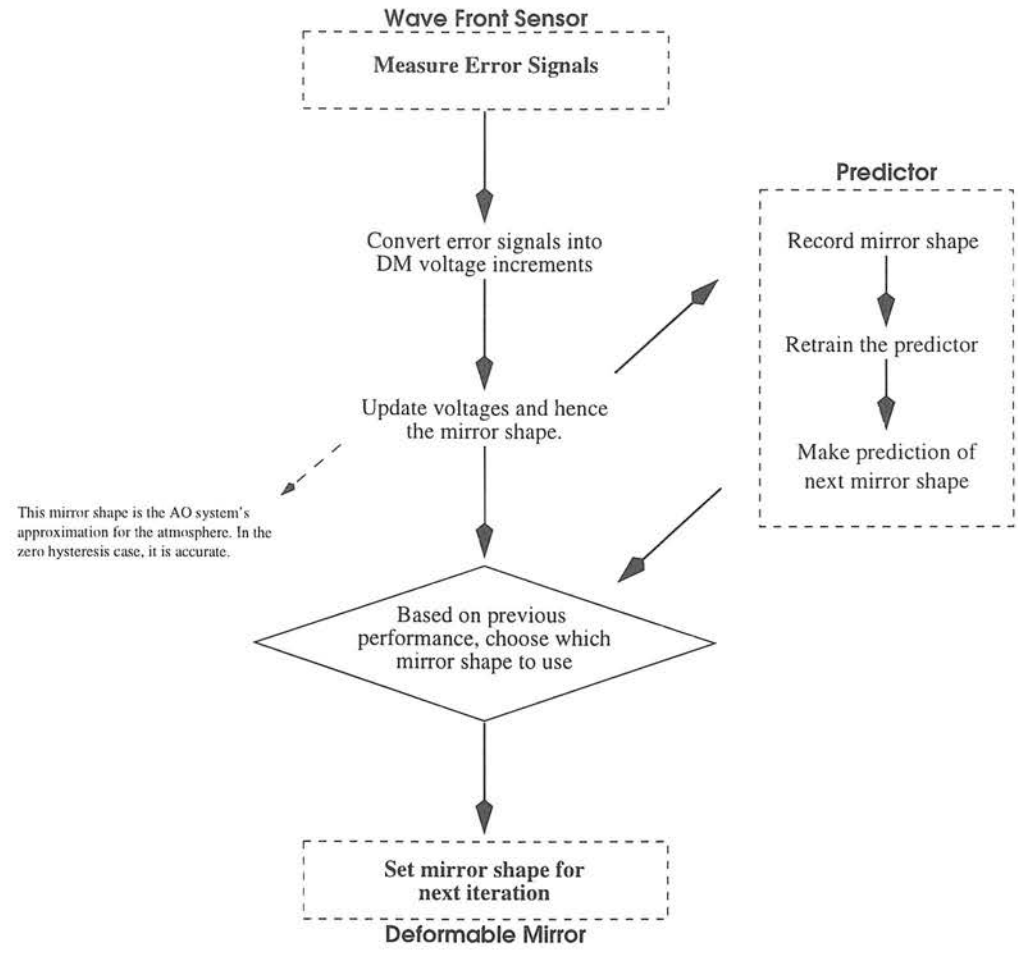


Figure 3.9: Schematic of the predictor implementation. The vertical path minus the rhombus is the standard AO system control loop. The addition is the predictor path and choice between the two predictions.

improvement, we describe the predictor as "well trained". In the neural network case, if it is "well trained", the number of training epochs per AO system iteration is reduced to minimise run–time.

The AO system now has two predictions for what the atmosphere will do next: that of the unaided AO system and that of the predictor. The crucial element of this scheme is that the system is able to test which prediction has been performing better in the most recent past and select appropriately. A schematic diagram of

this implementation is shown in figure 3.9.

The test is performed by comparing the variance of the error signal measured by the wavefront sensor with that that would have been obtained using the other predicted mirror shape. This is possible as the difference in the predicted mirror shapes is known. Note that it would not be possible to perform this test if the mirror suffered hysteretic effects.

The default decision is to use the *best is last* AO system shape and there are three conditions which must be satisfied before the prediction is used. Most important first:

1. the prediction in the previous AO system iteration must have performed better than the unaided AO system,
2. the predictor must be performing reasonably well over the training set and
3. the correction error signal $E(\underline{r})$ is small.

This decision stage is essential as the atmosphere can evolve on a short timescale causing the predictor to become out of date. The predictor does not always outperform the *best is last* prediction from the AO system so it is essential that the control loop can determine which estimate to use.

Unfortunately, there is still a lag in the system. It is now a lag in determining whether or not to use the predicted mirror shape. It is conceivable that there

are many single, isolated examples of the predictor outperforming the unaided AO system. However, because the choice of mirror shape to use is based on the performance in the previous iteration, it is possible that the overall performance of the system will suffer. An overall gain in system performance requires that there are clumps of AO system iterations for which the predicted mirror shape outperforms the unaided AO system.

3.10.1 Routines

IDL comes supplied with a set of routines which perform time series prediction using linear predictors. These routines, TS_COEF, TS_DIFF, TS_FCAST and TS_SMOOTH, were used throughout this work to investigate the linear predictor case (RSI, 1994). It was necessary to implement neural network algorithms, first as a stand–alone programme for testing purposes and later combined with the AO simulation.

3.11 Rigaut's AO Model

The AO model which has been used throughout this work was kindly supplied by Francois Rigaut of the Canada France Hawaii Telescope. The model is implemented in the Interactive Data Language (IDL) which is similar to FORTRAN in syntax, object oriented in its approach and comes supplied with a large range of in–built libraries to perform such tasks as fast–Fourier transforms and sorting.

The astronomical community has also made use of the language and there exists a large collection of routines specific to astronomical applications.

Rigaut's model also has a large number of parameters which one can vary. Fortunately, a useful set can be found in the paper in which his model is described (Rigaut et al., 1997). The parameters are assigned in a configuration file and specify such things as the telescope geometry, WFS geometry, DM configuration as well as the integration time per AO system iteration and the control loop gain. In all there are around 70 parameters which can be varied. The manual which accompanies the model contains a full description of each parameter.

The model comes supplied with 3 modelled atmospheric phase screens. The heights and horizontal speeds of these screens are specified in the parameter file. Investigations of the structure function of these screens showed them to be reasonable approximations of Kolmogorov turbulence as they exhibited the $r^{\frac{5}{3}}$ structure function.

Before the model can be run, it is necessary to initialise the WFS and DM by converting the file parameters into model representations of the WFS and DM.

For the WFS, this involves projecting the appropriate sub–aperture pattern onto the simulated telescope pupil. The illumination of each sub–aperture is determined and if one is very poorly illuminated then the corresponding image is ignored in subsequent steps.

The DM step is slightly more complicated. The positions of the actuators are defined on a regular grid. For the Shack–Hartmann case, there should be one more actuator in each direction than there are sub–apertures so that the actuators are situated at the corners of the sub–apertures. The influence of each actuator on the surface figure of the mirror to unit voltage is obtained through applying a model influence function. If there is a separate tip/tilt mirror, then this is treated as two additional actuators on the DM with their own specific influence function. These influence functions are then used in a SVD technique to find the normal modes of the mirror. As noted above, the mirror shape can be decomposed into these modes, the coefficients being the data used in the predictors.

The modelled mirror is of stacked actuator continuous faceplate design. As implemented, the DM has zero hysteresis and is therefore ideal for use in the prediction scheme discussed above.

3.12 Summary

In this chapter, the problem of atmospheric turbulence in astronomical observations has been discussed. The state–of–the–art solution, adaptive optics, has been described and the limitations of such a system outlined. A method of improving the performance of an AO system through extra software in the control loop to overcome the time lag between measurement and correction has been proposed. The next chapter details the experiments which have been performed and the

results obtained with the modified adaptive optics model.

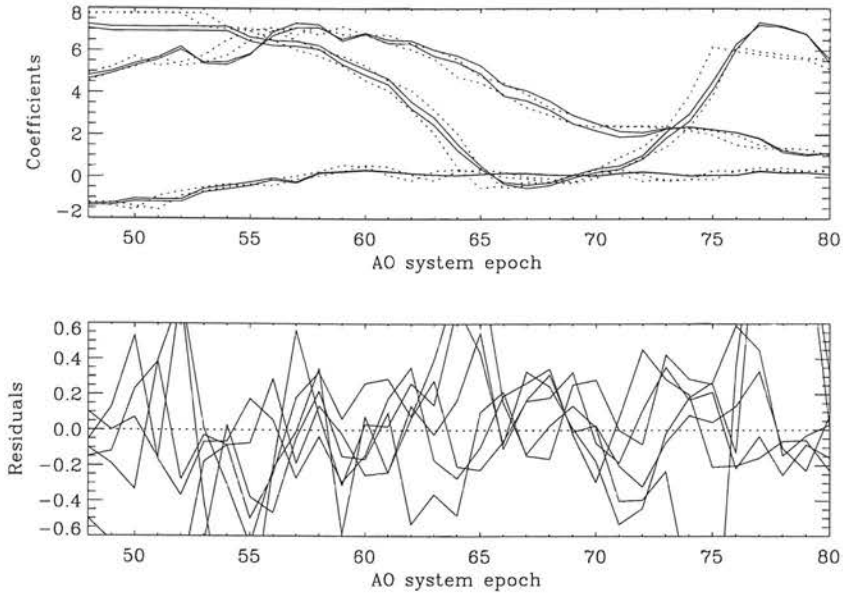
Chapter 4

Prediction Results

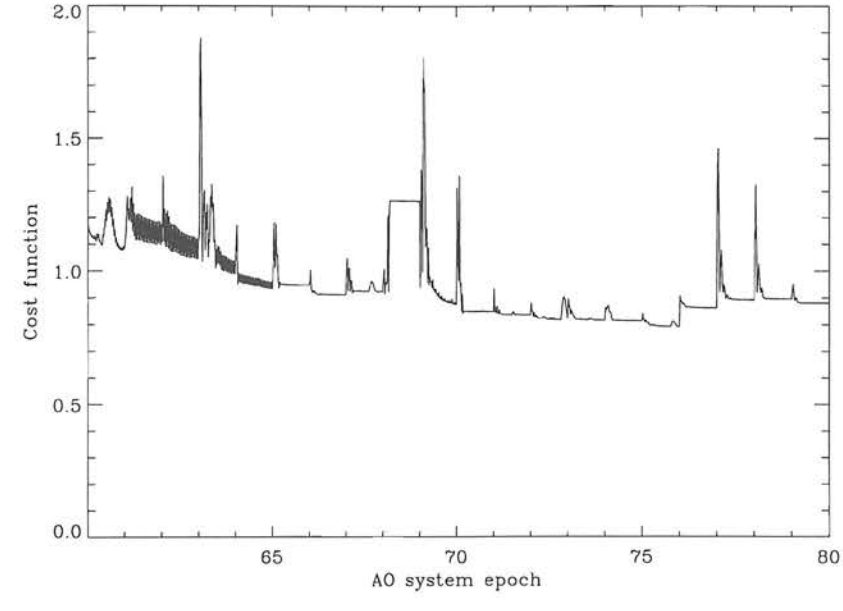
This chapter details a series of experiments, performed to investigate the behaviour of an AO system with an additional predictive element in the control loop. The aim of the predictor is to overcome the time lag between the measurement of a distortion and correction of that distortion with the deformable mirror. Eliminating the time lag through prediction should lead to an improvement in Strehl when compared with the non–predictive case. This improvement in performance will allow increased sky–coverage for AO systems.

4.1 Behaviour of the Predictors

Figure 4.1 shows a snapshot of the behaviour of a predictor during operation of the model. Figure 4.1(a) top shows the direct comparison of the values outputted by the predictor with those required by the training set. The predictor can determine the general functional form quite well, though there are significant



(a) Comparison of the learning set with the predictors representation. Top: a direct comparison of the coefficients, bottom: residuals after subtraction of the required datavalues. The dotted line shows the required datavalues, the solid line shows the output from the predictor.



(b) Evolution of the cost function through AO system iteration. When the function is below 1.0, the predictor is outperforming the unaided AO system over the training set.

Figure 4.1: A snapshot of the predictor characteristics.

departures apparent in figure 4.1(a) bottom. There are two traces for each of three coefficients as the predictor is a neural network being trained on the vector problem — section 3.9.5.

With each iteration of the AO system, the oldest member of the training set is discarded to make way for the most recent complete example of inputs and outputs. The new piece of data usually causes the performance of the system, measured by the cost function, to worsen before retaining of the predictor takes place. This is apparent in figure 4.1 (b) where the spikes indicate the updating of the training set.

The predictor in this example is a neural network, and there are 50 training epochs per AO system iteration — it is clear that there are usually enough training epochs per AO system iteration for the neural network to retrain before the next update to the training set. The cost function in this case is a comparison of the predictor performance with the unaided AO system performance. When the function is below 1.0, the predictor is outperforming the unaided AO system over the training set.

4.2 Proving Trials

The first goal was to determine if it is possible to overcome the time lag between measurement and correction using a predictor in the control loop. A direct result

of the prediction should be an overall improvement in system performance as measured by the Strehl ratio of the long–exposure science image, built up in each run of the model AO system (section 3.6.3). The relative merit of linear and neural network predictors was also of interest.

To minimise run–time, a low–order system of a 4×4 subaperture Shack–Hartmann sensor mated with a 5×5 actuator deformable mirror was modelled. The system was simulated to be working on the William Hershel Telescope (WHT).

In this configuration, the corner subapertures of the WFS are poorly illuminated and ignored by the model. The corner actuators of the DM have no effect beyond the corner subapertures of the WFS, so were also ignored. This left 21 actuators, which coupled with a separate tip/tilt mirror gave the mirror 23 degrees of freedom. In principle, this gave the system 23 degrees of freedom, or modes of correction. However, in practise there are certain modes of the DM which the WFS cannot measure. These modes are the null–space found during the inversion of the interaction matrix (section 3.6.2) and are discarded when creating the control matrix.

The number of modes of correction actually used – the degree of correction — was further reduced during the matrix inversion stage to leave only the 6 lowest order system modes. The object was to create a very simple model system.

The number of modes of the mirror was also 23 given the 23 degrees of freedom

provided by the actuators — the shape of these modes is determined by a routine supplied with the model. To match the degree of correction, the coefficients of the 6 lowest order mirror modes were recorded into the training set for the predictor.

The number of records of the AO system shown to the predictor inputs was arbitrarily chosen as 4 and in the neural network case, the predictor was trained with the two element vector problem. The number of hidden layer nodes in the single hidden layer was set equal to the number of outputs. This gave a network of 24 inputs, 12 hidden layer nodes and 12 outputs. The number of coefficients, d_j, in the linear predictor was set at 24 to keep the characteristics of the two predictors as equivalent as possible.

The relevant AO system parameters used in this case are detailed in table 4.1. The tabulated heights of the turbulent layers have no effect on the performance as only the on–axis case is being considered. The value of the sky background is unrealistically faint. This is rectified at a later stage.

An investigation of Strehl with integration time was performed for three cases, unaided AO, neural network prediction and linear prediction. The number of AO system iterations were varied as appropriate to keep the total system time constant. This meant that the same piece of modelled atmosphere was viewed in each case, enabling as fair a comparison as possible to be made. The results are graphed in figure 4.2.

AO System Parameter	Value
WFS wavelength:	$0.7\mu m$
Science wavelength:	$1.6\mu m$
Total time of run:	$2.75s$
Servo warm–up time:	$0.25s$
Length of training set:	$0.5s$
Telescope diameter:	$4.2m$
Strength of phase screen 1:	0.2
Strength of phase screen 2:	0.6
Strength of phase screen 3:	0.2
Speed of screen 1 (+ve x–dir^n):	$15ms^{-1}$
Speed of screen 2 (+ve y–dir^n):	$20ms^{-1}$
Speed of screen 3 (+ve x–dir^n):	$29ms^{-1}$
Height of screen 1:	$10m$
Height of screen 2:	$1km$
Height of screen 3:	$10km$
L_0 of each screen:	∞m
Guide star magnitude at WFS λ:	15.5 magnitude
Sky background at WFS λ:	80.3 magnitude arcsec^{-2}
$r_0(0.55\mu m)$:	$15.7cm$
CCD read noise:	$3e^-$

Table 4.1: AO model parameters used to create figure 4.2. The relative heights of the phase screens are for completeness as they have no effect in the modelled on–axis case.

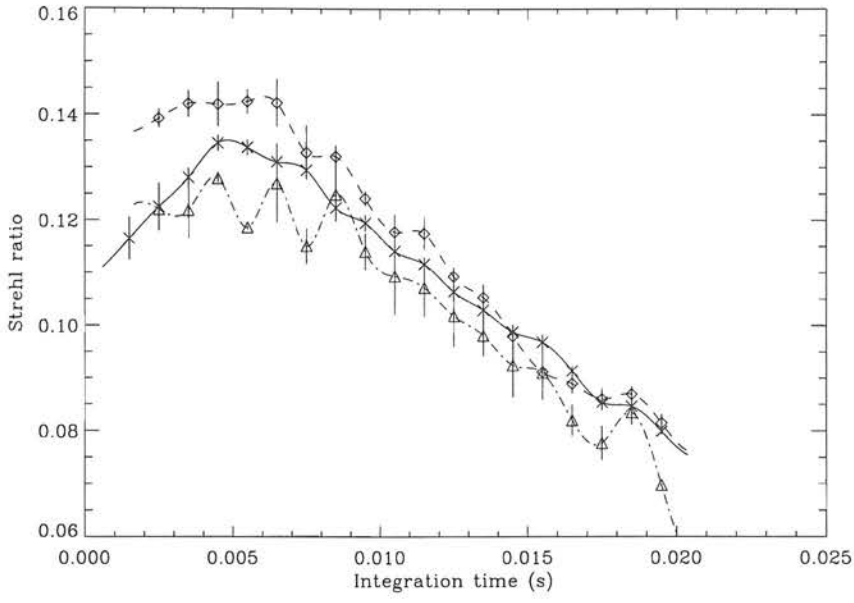


Figure 4.2: First comparison of AO system performance with and without predictors for parameters tabulated in table 4.1:

-×- no prediction,

-◇- neural network and

-△- linear predictor.

The solid line shows the unaided AO system. The behaviour outlined in section 3.5.5 is apparent. As the integration time per AO system step increases, there is an increase in performance as signal–to–noise improves. The increased decorrelation of the atmosphere causes a turn–over in performance at around $5ms$ and the performance drops off steadily as the integration time, and hence atmosphere decorrelation increases.

The diamonds show the performance with the option to use the NN prediction present in the simulation. It is clear that there is benefit to be obtained from the NN, as the peak performance has been improved, and the peak occurs at a noticibly longer integration time, of around 6.5ms.

The triangles show the performance with the option to use the LP prediction switched on. This line is consistently below that of the no prediction case, and shows that the option to use the LP was actually making performance worse.

4.2.1 Explanation of Poor LP Performance

Of interest is why there is a degradation of performance with the LP. One would naively expect that the base performance of the AO system without prediction would at least be maintained. Visual observation of the indicator of whether prediction was being used showed that in the NN case, clumps of 4 or 5 consecutive predictions were observed. Conversely, in the LP case isolated predictions were far more common. A more detailed experiment of this type was performed at a

later date and this is discussed in section 4.5.

The reason why isolated single instances of using the prediction are a bad thing is due to the lag in switching the prediction in and out of the control of the mirror. For the predictor to be switched in, the performance of the predictor in the *previous* iteration of the AO system is the main piece of data considered. With no prediction being used, if the previous prediction was good in that the performance of the system would have been better using it, then the predictor is switched in in the current iteration. It is perfectly possible that the predictor then makes a bad prediction. There is no way of determining if this is what will happen before it does. The system is constrained to use the bad prediction and can only switch the predictor out after the test in the next iteration when the prediction has proved to be bad.

Therefore, there is always a performance hit when the predictor is switched out. If the number of consecutive predictions is 1, then there will be a performance hit. If the number is 2 then there is a benefit for the first of the pair and a hit for the second of the pair which should roughly cancel. It is only when we get to clumps of 3 or greater that benefits start to be obtained when using the predictor.

Given that the LP gave worse results than no prediction, and that the NN was giving improved performance led, at this stage, to the NN being chosen for further investigations.

4.3 Practical Considerations

The computing power required to train a NN is roughly proportional to the square of the number of inputs. For a more realistic AO system which might have around 70 mirror modes, a NN characterising and predicting such a large number may not be feasible given the available computing power. Therefore, an experiment was performed to discover whether benefit would be obtained from predicting only the lowest mirror modes whilst the AO system was correcting a significantly greater number.

The previous parameters were used with the modification that the degree of correction of the AO system was now 16 modes, with the degree of prediction remaining at 6 modes. The results are graphed in figure 4.3.

This investigation proved a disappointment as the expected benefit of predicting the lowest modes did not appear. To try to understand this result, further tests were performed keeping the integration time and magnitude of the guide star constant, but varying the NN in two ways. Firstly, only one mode was predicted with the mode of choice being varied — single mode prediction. Secondly, each mode up to and including that specified were inputted to the NN – cumulative mode prediction. Figure 4.4 shows the difference between the two cases.

Figure 4.5 shows the results of the run. The single mode curve (crosses) shows that there is a clear benefit to be had by predicting most of the modes. What is

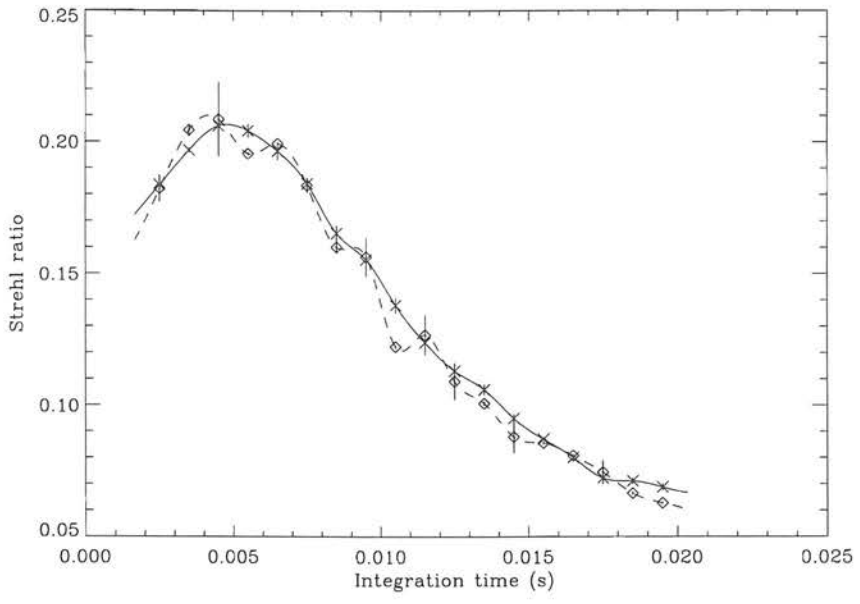


Figure 4.3: Further comparison of AO system performance with and without predictors: $\rightarrow\!\!\!\!\!\!\times\!\!\!\!\!\!\leftarrow$ no prediction and $-\!\!\diamond\!\!-$ neural network.

disappointing however is that the performance of the cumulative mode prediction falls off after a few modes. It does not, as one might expect, accumulate the benefits to be had from each mode included in the predictor.

This suggests that the cumulative method has a flaw. For there to be an accumulation of the benefits from the prediction for each individual mode, the cumulative method requires that the predictions of each mode are good or bad in a coincident fashion, *i.e.* all good or all bad. This is highly unlikely, and while there may be a benefit from predicting some modes, this is cancelled out by the other modes for which there is a performance hit.

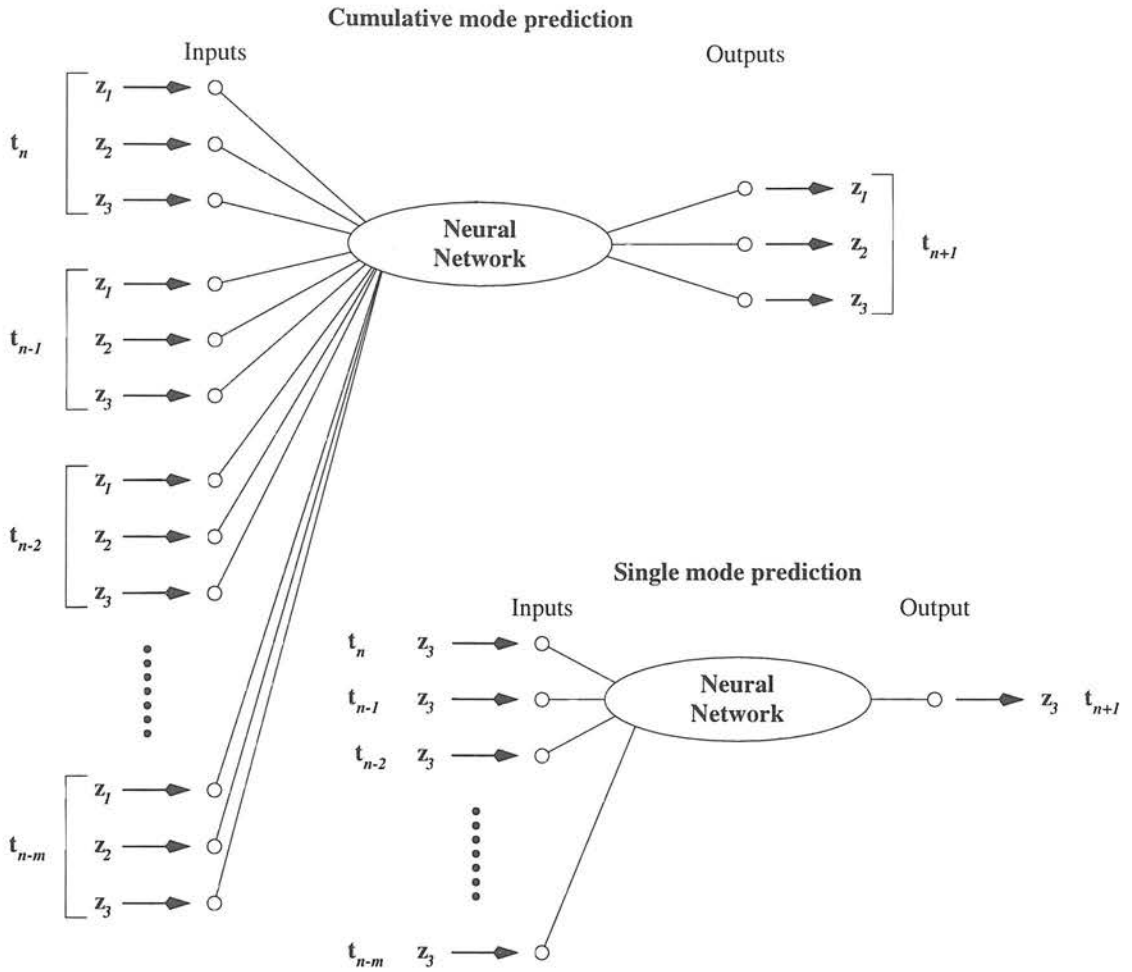


Figure 4.4: Cumulative vs. single mode prediction. The window size for each network is the same, the modes which are inputted and predicted are different.

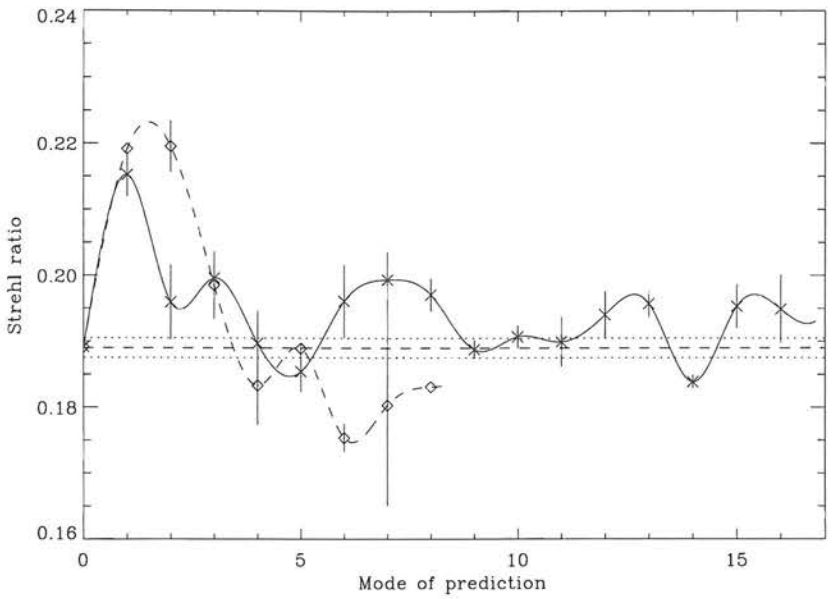


Figure 4.5: Comparison of prediction with AO system correcting 16 modes: Horizontal lines: no prediction case, the dashed line is mean performance with the dotted lines showing ± one standard deviation,

-×- predicting only specified mode and

-◇- predicting up to and including specified mode.

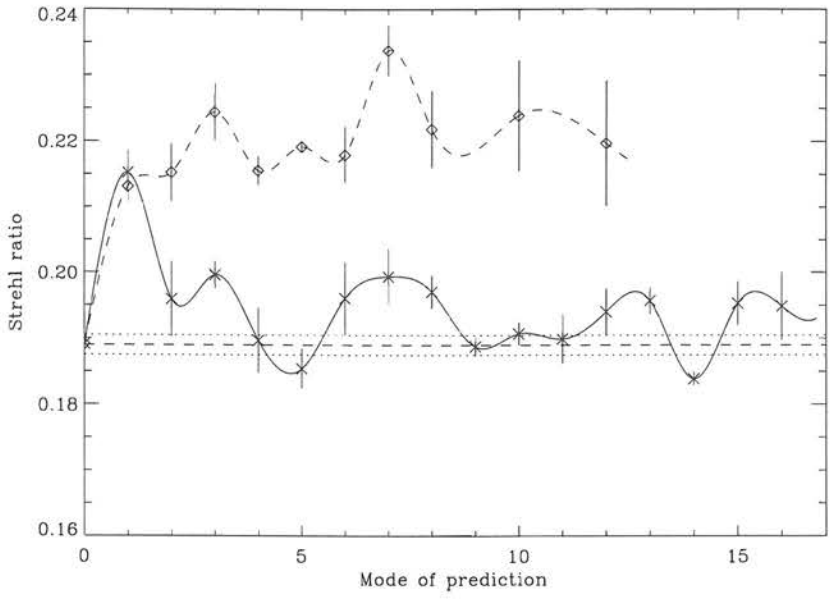


Figure 4.6: Comparison of prediction with AO system correcting 16 modes:
Horizontal lines: no prediction case, the dashed line is mean performance with the dotted lines showing ± one standard deviation,
-×- one neural network predicting only specified mode and
-◇- parallel neural networks, their number equalling the specified mode.

4.4 Parallel Predictors

The above argument suggests that instead of one large NN when we wish to predict many modes, we should use many single mode predictors in parallel. This approach is also used by the ONERA group (Dessenne et al., 1998). The full benefit of prediction should be obtained when the performance of each individual network is tested in the same fashion as before and only those predictions which have shown to be good are selected. Otherwise, the values for the modes suggested by the unaided AO system are used.

To test this assertion, the cumulative case was re–run, but with parallel single

mode NN's rather than the one large NN that was used in the cumulative case. The expected accumulation of the performance available from the single modes is shown in figure 4.6 confirming that many parallel predictors is the method of choice. Note that the individual neural networks used are now very simple, and therefore easy to implement and relatively quick to train. The input layer is 4 nodes, the hidden layer is 2 nodes and the output layer is 2 nodes (the vector problem), with only one being used when a prediction is being made.

Also confirmed by this graph is the suggestion from section 4.3 that predicting only the lowest mirror modes while the AO system is correcting a greater number is possible. Thus, if there are limits on the computing power available for prediction, there will still be a benefit from predicting only the lowest modes.

4.4.1 Linear Predictors Revisited

Having established that the correct implementation of prediction is with many parallel single mode predictors, it is worthwhile checking to see if the LP case was being hampered by not using this implementation. A similar experiment to that used to create figure 4.2 was re–run with the parallel prediction case. In this example, the magnitude of the guide star was set at 15.0.

Total system time was again $2.75s$, with performance being averaged over the last $1.75s$. The servo warm–up period was $0.25s$ and the training set length was $0.5s$, giving $0.25s$ of system time for the predictor to converge before the measurement

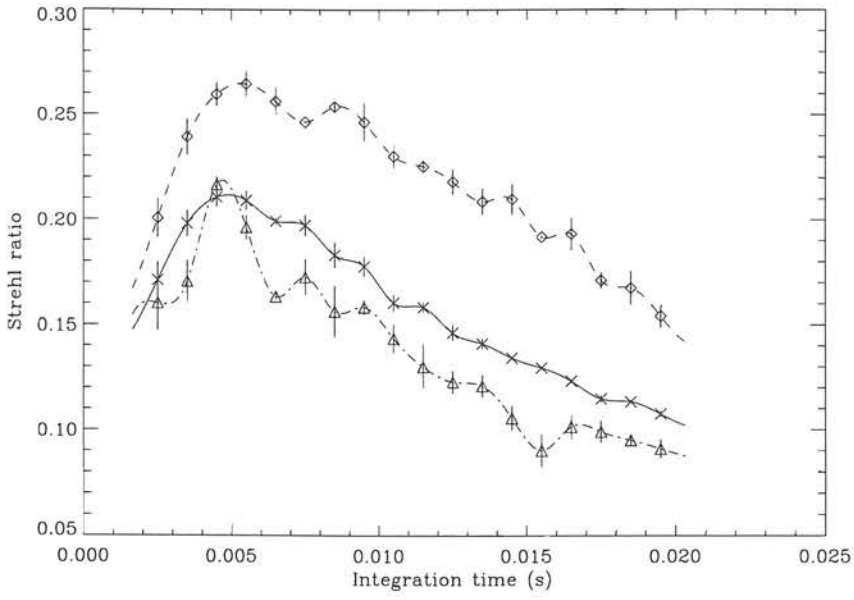


Figure 4.7: Comparison of performance with parallel predictors:
-×- no prediction,
-◇- neural networks and
-△- linear predictors.

of the performance began. The results for this run are shown in figure 4.7.

It is clear that the LP is still performing worse than the unaided AO system. However, as mentioned earlier, this was for an unrealistically low sky brightness. Re–running the model with a more suitable value of 20.3 magnitude arcsec^{-2} (Rigaut et al., 1997) shows that the LP case becomes equivalent to the no–prediction AO case.

The decrease in signal to noise caused by the increase in sky brightness has two effects. Firstly, the peak performance suffers due to the decrease in signal to noise as one would expect. Secondly, it forces an increase in the total system time to $10s$ with performance data being collected for $9s$. This is required to provide

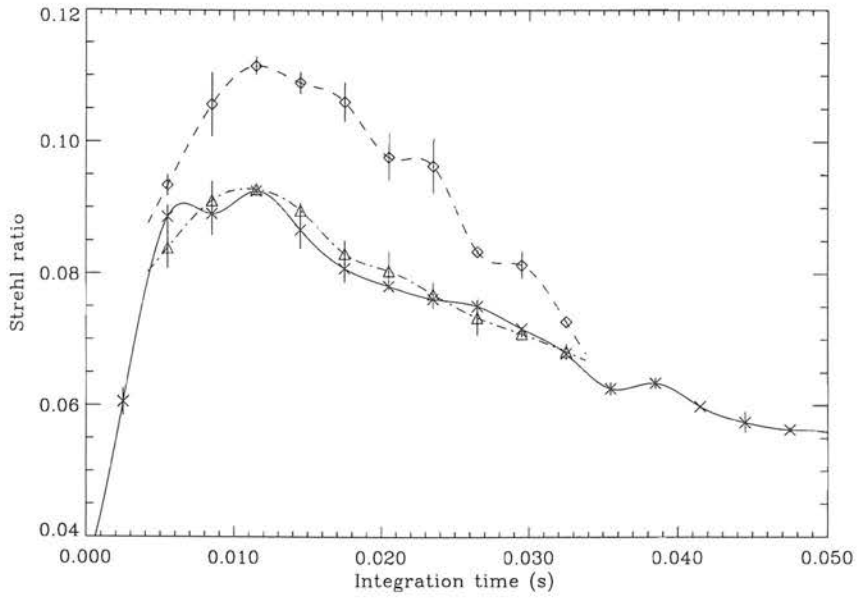


Figure 4.8: Re–run of figure 4.7 with realistic sky brightness:
–×– no prediction,
–◇– neural networks and
–△– linear predictors.

enough samples of the noisier measurements to overcome the noise and resolve the differences between the cases. This does mean that the atmosphere sampled is different to that in figure 4.7, however, the first $2.75s$ are the same. The length of the training set and servo warm–up periods remain the same. The NN still has a clear advantage as shown in figure 4.8.

4.5 Clumps of Prediction

As mentioned earlier in section 4.2.1, the number of consecutive AO system iterations which use prediction has an effect on the increase in performance of the system. The runs for the predictors at the integration times $0.015s$ in figure 4.7

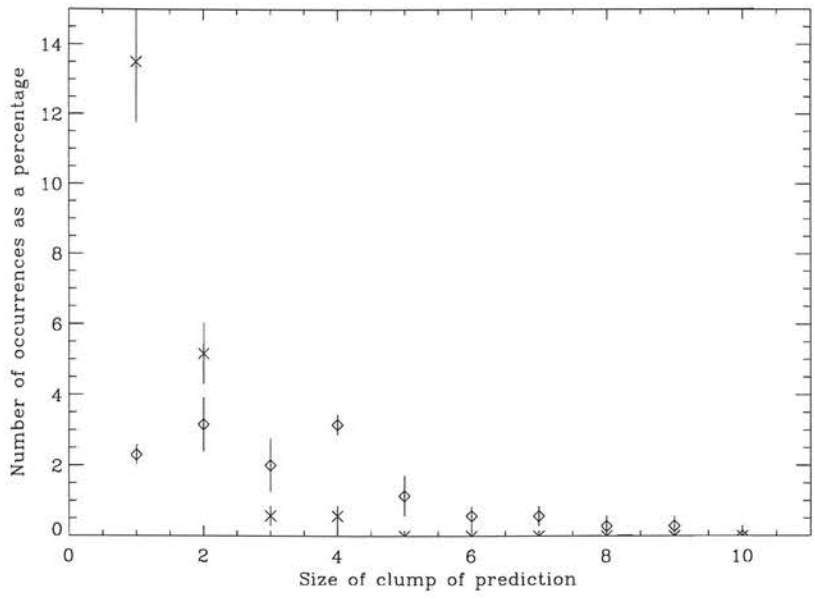


Figure 4.9: Histogram of size of clumps for integration time of $0.015s$ in figure 4.7. Crosses: linear predictors, diamonds: neural networks. Iterations not using prediction: LP = 72%, NN = 55%.

and $0.0205s$ in figure 4.8 were repeated, recording the number of occurrences of the clumps of prediction of each size. These histograms were then normalised by the number of iterations in the AO loop, creating comparable histograms of the number of consecutive AO system iterations which used prediction. The histograms for the mirror "tip"mode are figures 4.9 and 4.10 for the two cases respectively.

Of particular interest is the relative strength of the number of times a clump of size one occurs. As noted earlier, due to the lag in switching out the prediction, this case will always lead to a loss in performance. In figure 4.10, the relative strength of the *Size of clump* = 1 LP point has been reduced dramatically when compared with figure 4.9. At the same time, the performance of the LP has

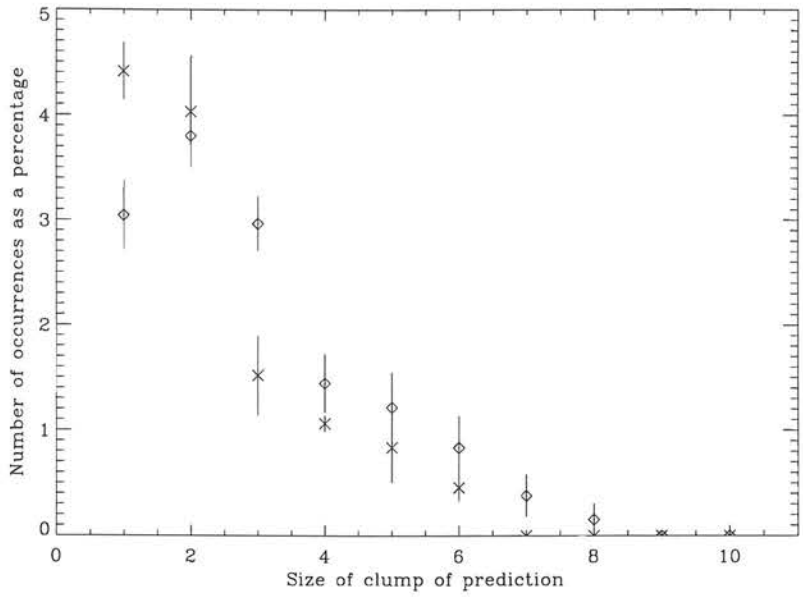


Figure 4.10: Histogram of size of clumps for integration time of $0.0205s$ in figure 4.8. Crosses: linear predictors, diamonds: neural networks. Iterations not using prediction: LP = 72%, NN = 60%.

improved from being below the unaided AO system to being equivalent.

The NN's on the other hand have relatively few examples of using single, isolated occurrences of prediction which would hit performance. For clumps of size 3 and greater, those which will always provide a boost to performance, there are always more NN examples.

4.6 Maximum Achievable Performance with Predictors

An interesting question is to compare the performance obtained with predictors with that that could be obtained if the predictors were performing perfectly, *i.e.*

removing all time lag in the control loop. A similar investigation to that in figure 4.6 was performed using the neural network predictors for the following cases:

1. The model was run with perfect prediction for each mode in turn showing the improvement available for each mode (diamonds) through removing the time lag.

2. The model was run with perfect prediction for up to and including the specified mode to show the accumulated improvement (crosses) through removing the time lag.

3. The model was run in the normal way with the neural networks option enabled (triangles), varying the number of modes predicted.

Note that this investigation is performed in realistic conditions with noise in the AO system.

The graph in figure 4.11 shows the expected performance of the perfect single mode and perfect accumulated prediction runs. What is of interest is the relative performance of the neural networks and the perfect accumulator. The graph shows that for this particular set of parameters, the parallel neural networks are performing very well when compared to perfect prediction. The curves cross at 3 modes of prediction, and for the 1 and 2 modes cases, the parallel neural networks are outperforming the "perfect prediction" case.

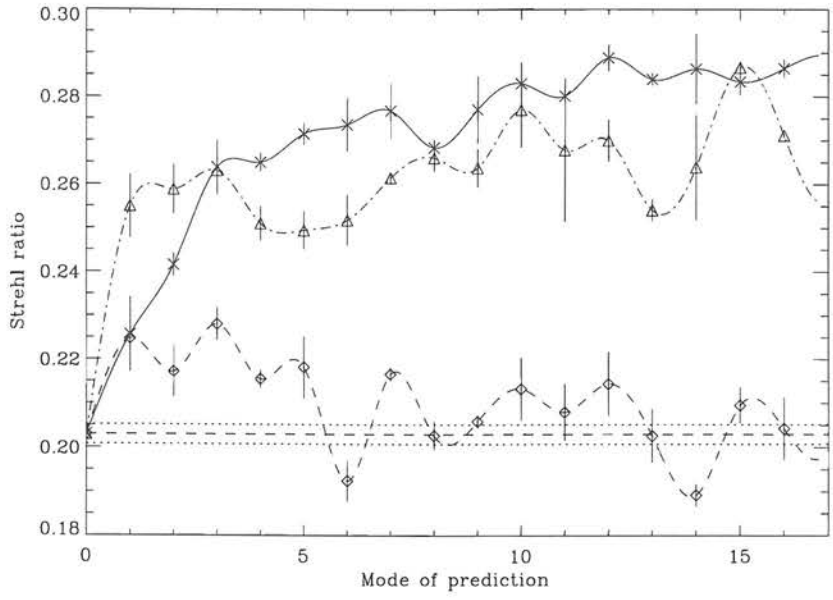


Figure 4.11: Comparison of neural networks and limiting prediction case:
Horizontal lines: no prediction case,
-×- Perfect accumulated prediction,
-◇- Perfect single mode prediction and
-△- Parallel neural networks.

This appears to be a contradiction, how can the predictor outperform the "perfect prediction" case?

A possible explanation is that the switches which choose between the unaided AO system modes and the predictor modes are responsible. The perfect prediction case has been created in the presence of noise, so the perfect predictions themselves, being what the system measures, are noisy. The tests performed to switch the various predictors in and out of the control loop choose the best performing set of modes. The best performers could be the least noisy predictions. Tip and tilt are the modes which are contradictory, and are the distortions with the greatest power (Noll, 1976). Any reduction in noise in these modes will have a

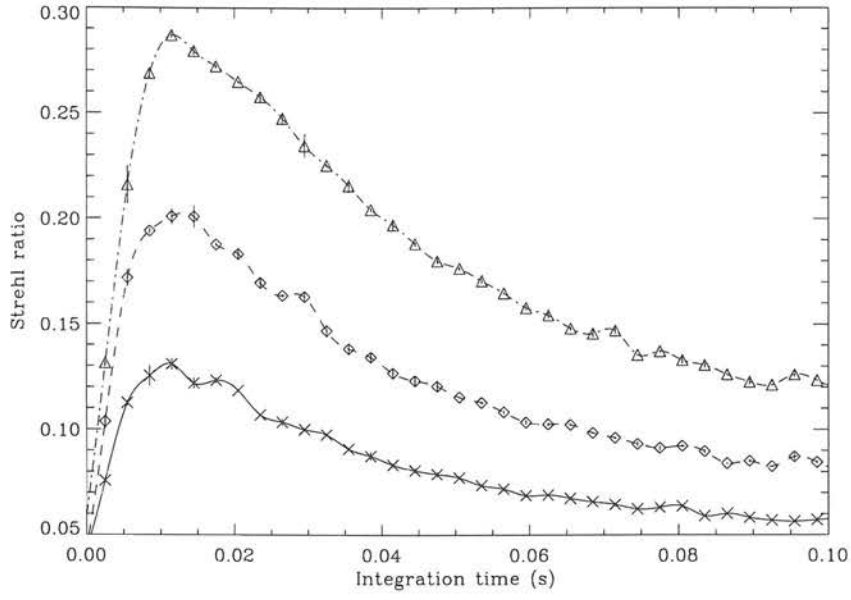
large effect on the system performance and could be enough to outperform the "perfect prediction" case.

Figure 4.11 can also be viewed as an optimisation graph for the degree of prediction. A balance has to be struck between the number of modes of prediction and run–time of the model. From figure 4.11, the degree of prediction is set at 3 modes, and this has been used in all subsequent runs of the model.

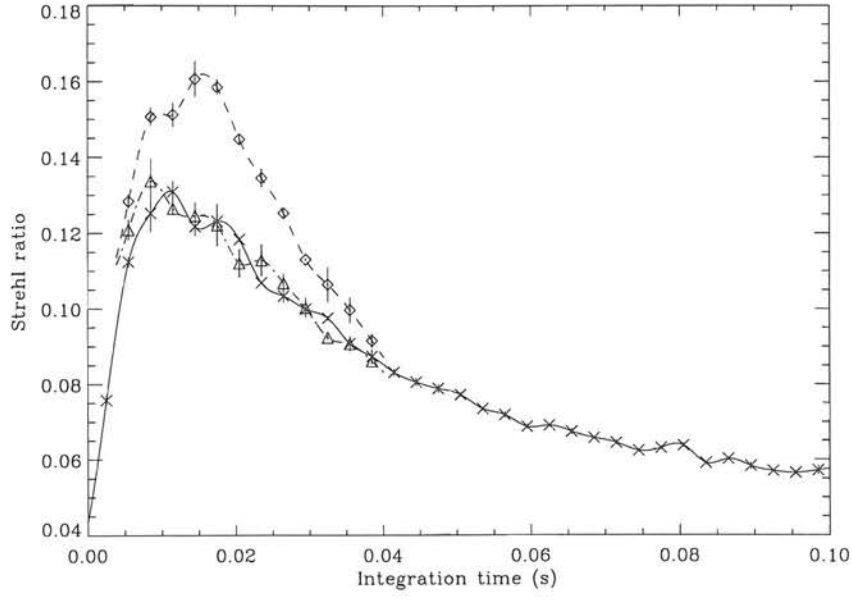
4.7 Variation of $r_0(0.55\mu m)$

As mentioned in section 3.8, previous work by Aitken and McGaughey suggests that the relative performance of neural networks and linear predictors changes as the coherence length of the atmosphere, $r_0(0.55\mu m)$, varies (Aitken and McGaughey, 1995). They state that for good seeing, LP's and NN's should give comparable performance whereas for poor seeing, the performance of NN's is greater than that of LP's.

The previous results have all been for an atmosphere amalgamated from 3 separate layers averaged so that $r_0(0.55\mu m) = 15.7cm$. To test whether the relative performance of NN's and LP's changed, a series of runs were performed for $r_0(0.55\mu m) = 20, 25$ and $30cm$ — these results are shown in figure 4.12. These plots show that for all the range of $r_0(0.55\mu m)$ the parallel neural networks outperform the parallel linear predictors, with the linear predictor case

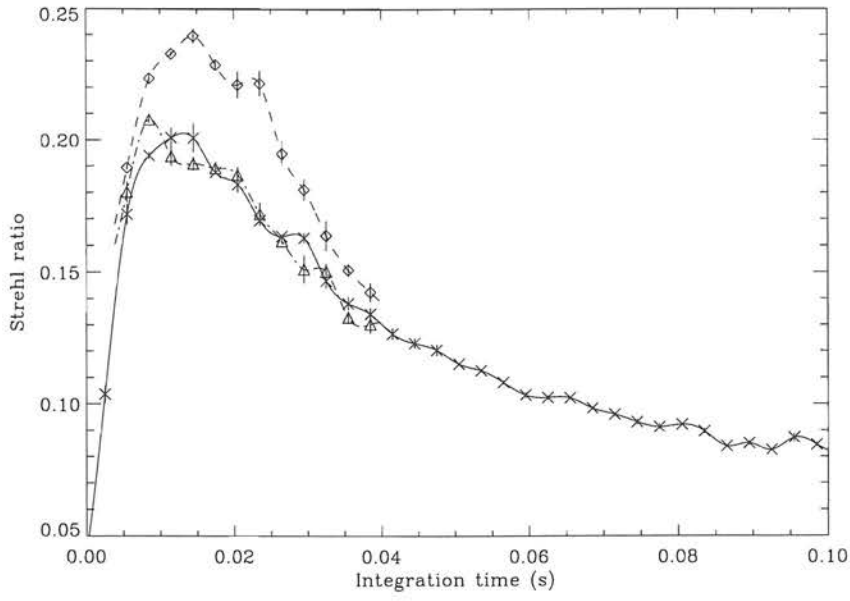


(a) Comparison of no–prediction AO system with r_0:
- $r_0(0.55\mu m) = 20cm$,
- $r_0(0.55\mu m) = 25cm$ and
- $r_0(0.55\mu m) = 30cm$.

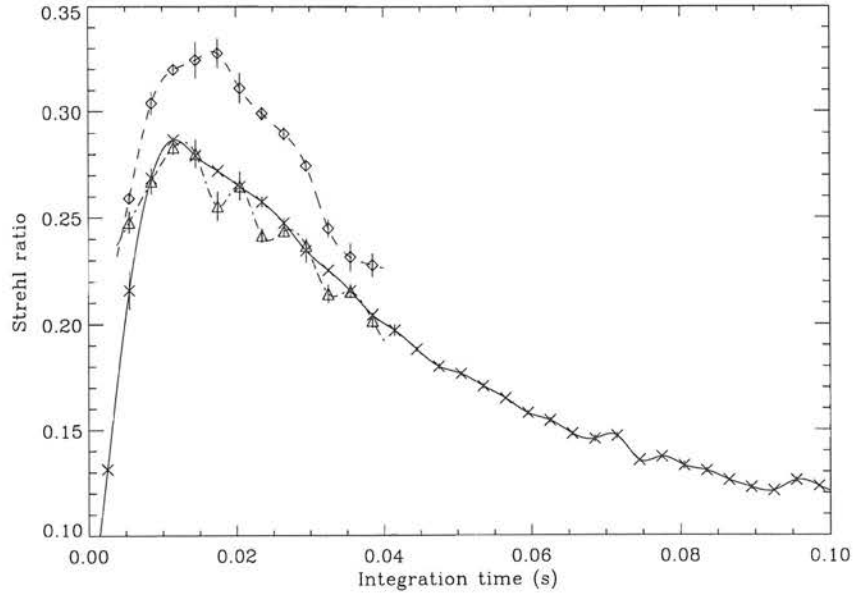


(b) No prediction, LP and NN performance with $r_0(0.55\mu m) = 20cm$:
- no prediction,
- neural networks and
- linear predictors.

Figure 4.12: Investigation of performance with r_0.



(c) As (b) with $r_0(0.55\mu m) = 25cm$.



(d) As (b) with $r_0(0.55\mu m) = 30cm$.

Figure 4.12: (*cont.*): Investigation of performance with r_0.

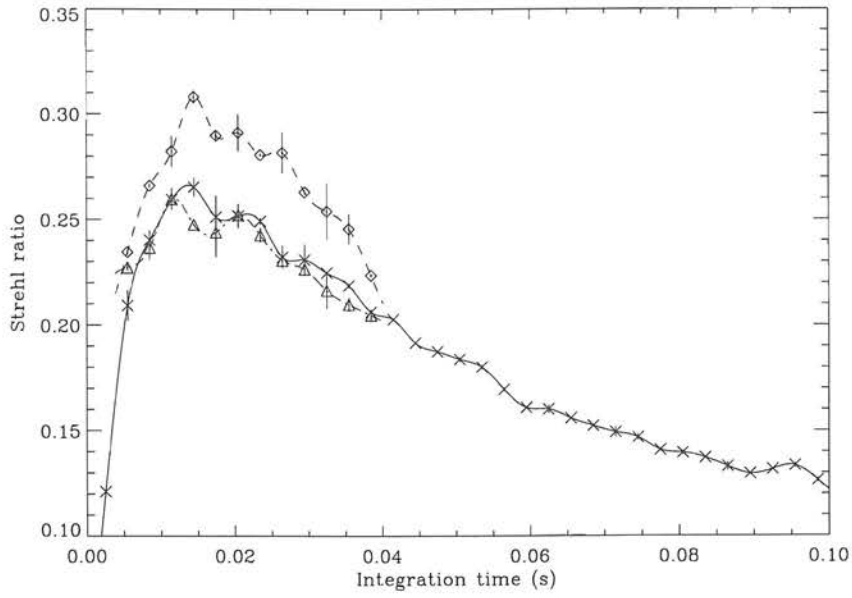
being equivalent to the unaided AO system.

4.8 Individual Atmospheres

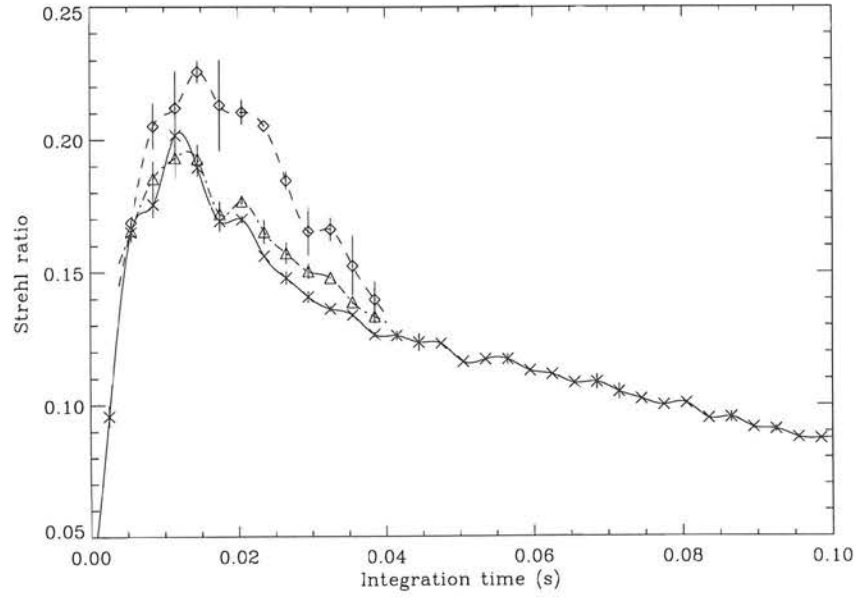
The previous runs of the model have all used the three supplied atmospheres in combination to provide a single, averaged atmosphere for the simulation. To obtain better statistics on the range of performance available from the AO system, the atmospheres are now presented to the model one at a time. The plots in figure 4.13 show the performance of the system to each atmosphere in turn. $r_0(0.55\mu m) = 25cm$ for these runs with the final plot, figure 4.13 (d), showing the performance averaged over the three atmospheres. These plots confirm that the neural networks still outperform the unaided AO system and the linear predictors.

4.9 Realistic Simulations

The previous simulations have been concerned with a simple 4×4 subaperture Shack–Hartmann system. More realistic systems such as the NAOMI system use 8×8 subapertures. Figure 4.14 shows the performance of such a system with the same atmospheres and guide star illumination as in figure 4.13. It is instructive that the performance is reduced significantly. This is due to the decrease in signal to noise in each subaperture as the amount of light in each has dropped by a factor of 4 — section 3.5.2. The system is therefore in the regime where suppression of higher order modes of correction is required to optimise the system's performance.

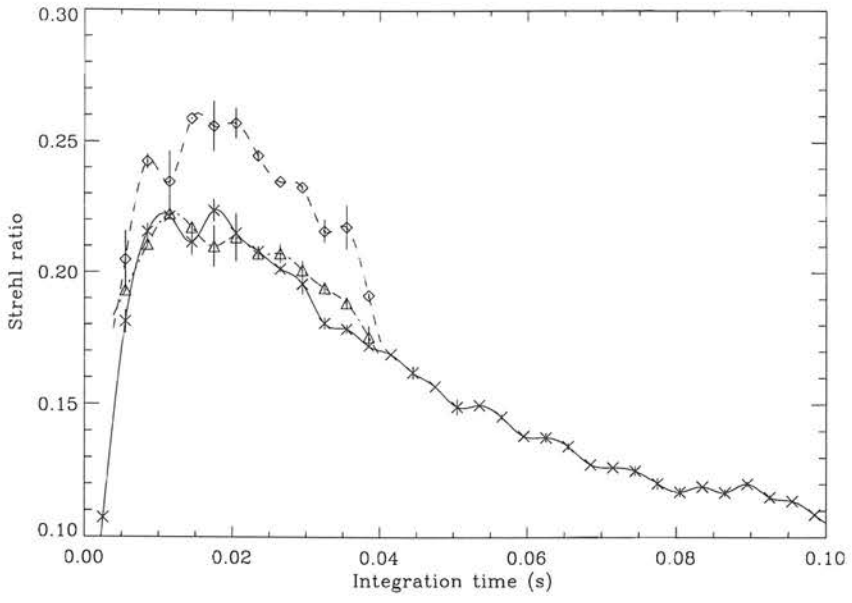


(a) Performance of system with atmosphere 1:
- ⟶ no prediction,
- -◇- neural networks and
- -△- linear predictors.

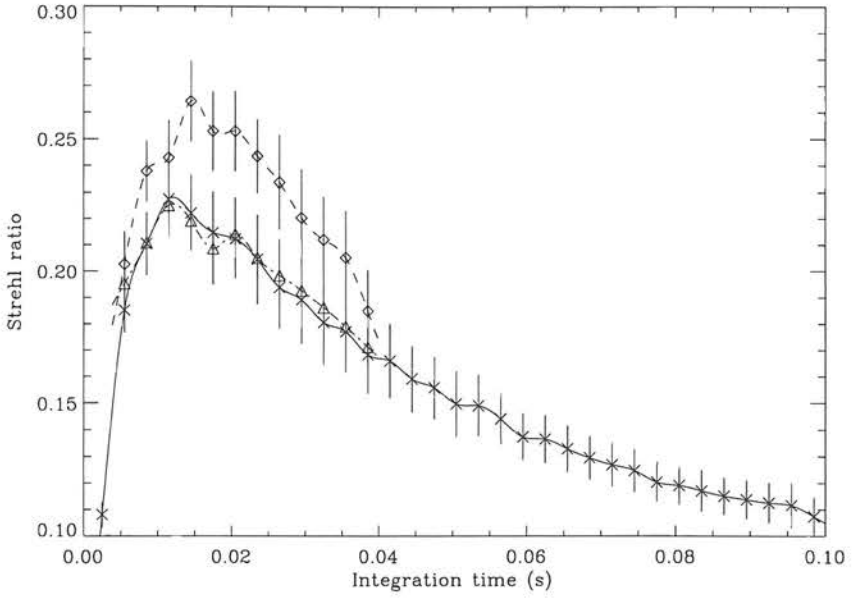


(b) As (a), but with atmosphere 2.

Figure 4.13: Investigation of 4×4 SH system performance with each phase screen.



(c) As (a), but with atmosphere 3.



(d) Performance averaged over the three atmospheres:
- -×- no prediction,
- -◇- neural networks and
- -△- linear predictors.

Figure 4.13: (*cont.*): Investigation of 4×4 SH system performance with each phase screen.

To save time, this has not been done in this example.

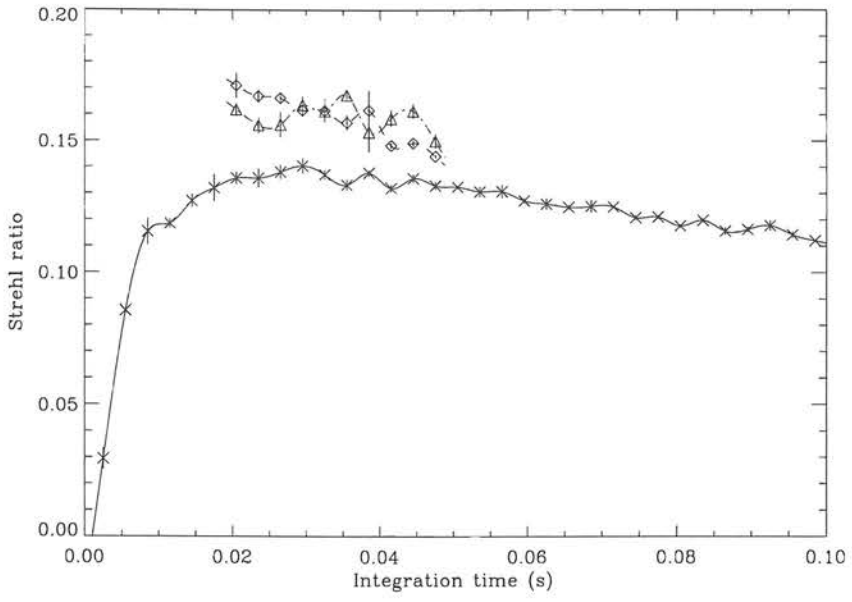
Figure 4.14 shows that the performance of the predictors is comparable for good seeing of $r_0(0.55\mu m) = 25cm$. As mentioned above, Aitken and McGaughey have suggested that for good seeing, the performance of the predictors should be comparable. Their proposition may be valid in the 8×8 subaperture case?

An investigation was performed with $r_0(0.55\mu m) = 15cm$ to test this hypothesis for the 8×8 subaperture case. If the proposition is valid, we expect that the neural networks should outperform the linear predictors. The results are in figure 4.15. The averaged results, figure 4.15 (d), show that the performance of the predictors is matched for $r_0(0.55\mu m) = 15cm$ as well as $25cm$.

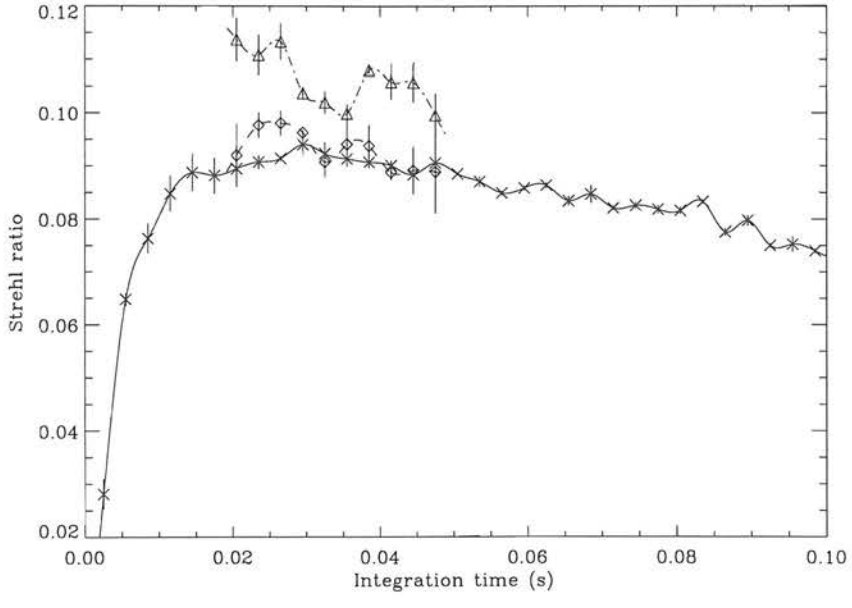
4.10 Performance Graph

The final goal of this work was to obtain a graph of the performance of the simulated NAOMI system with guide star magnitude, similar to that reported by Rigaut for his model of the proposed Gemini AO systems (Rigaut et al., 1997). The graph would compare unaided AO, AO with neural network predictors and AO with linear predictors.

As stated in section 3.10, the implementation of the predictive scheme is to be as straightforward as possible. Therefore, the case against which the prediction scheme should be compared is an optimised AO system. For this reason, good

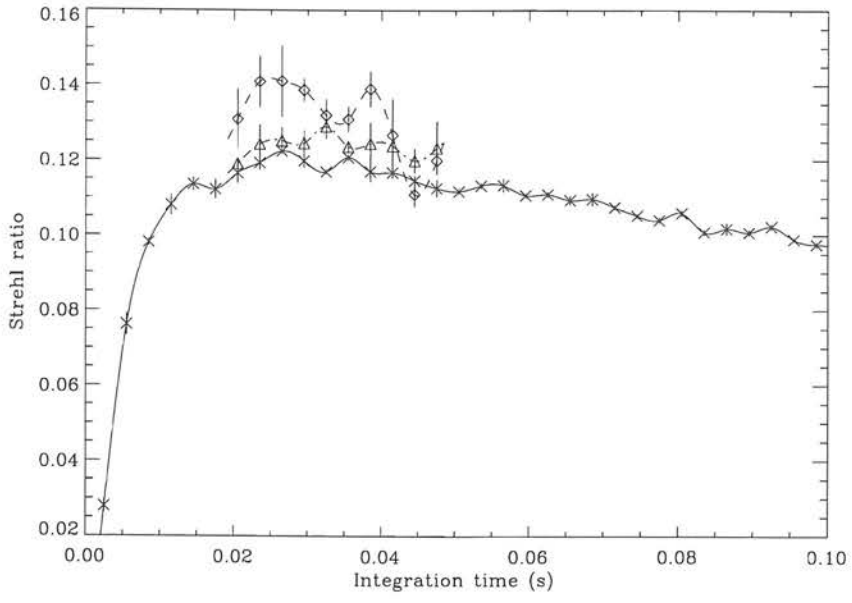


(a) Performance of system with atmosphere 1:
- -×- no prediction,
- -◇- neural networks and
- -△- linear predictors.

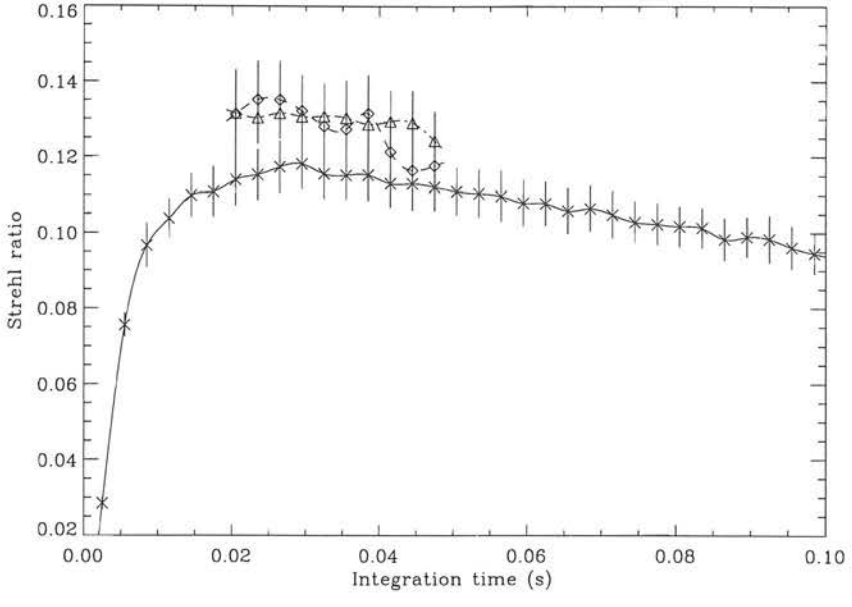


(b) As (a), but with atmosphere 2.

Figure 4.14: Investigation of 8×8 SH system performance with each phase screen with $r_0(0.55\mu m) = 25cm$.

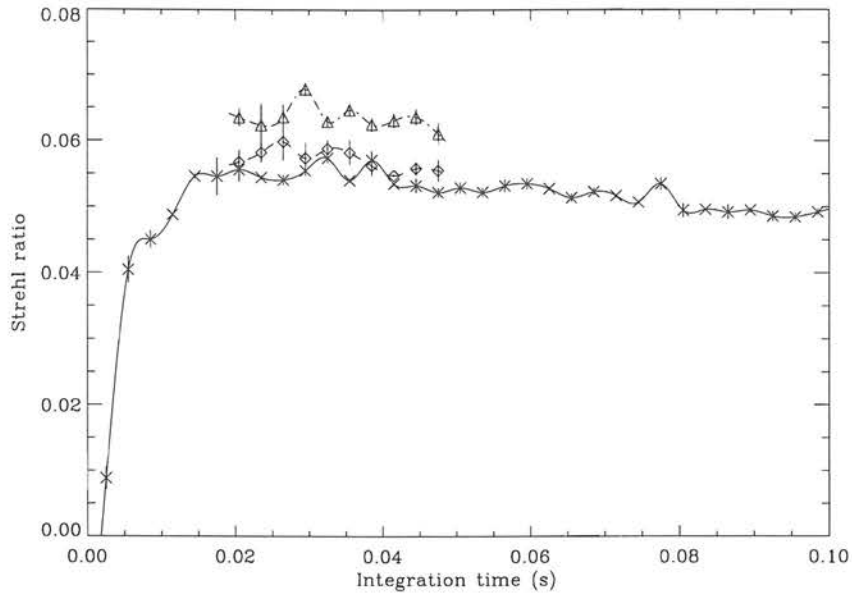


(c) As (a), but with atmosphere 3.

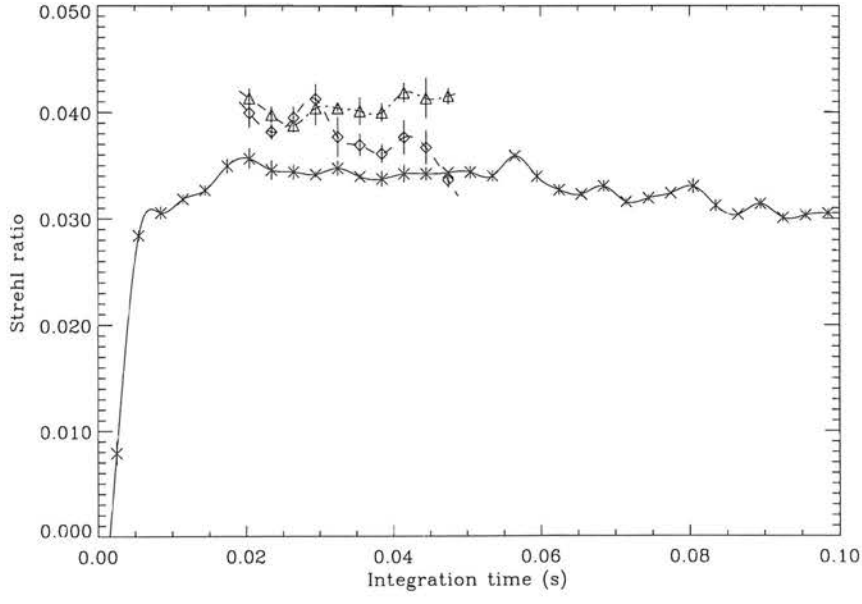


(d) Performance averaged over the three atmospheres:
- -×- no prediction,
- -◇- neural networks and
- -△- linear predictors.

Figure 4.14: (*cont.*): Investigation of 8×8 SH system performance with each phase screen with $r_0(0.55\mu m) = 25cm$.

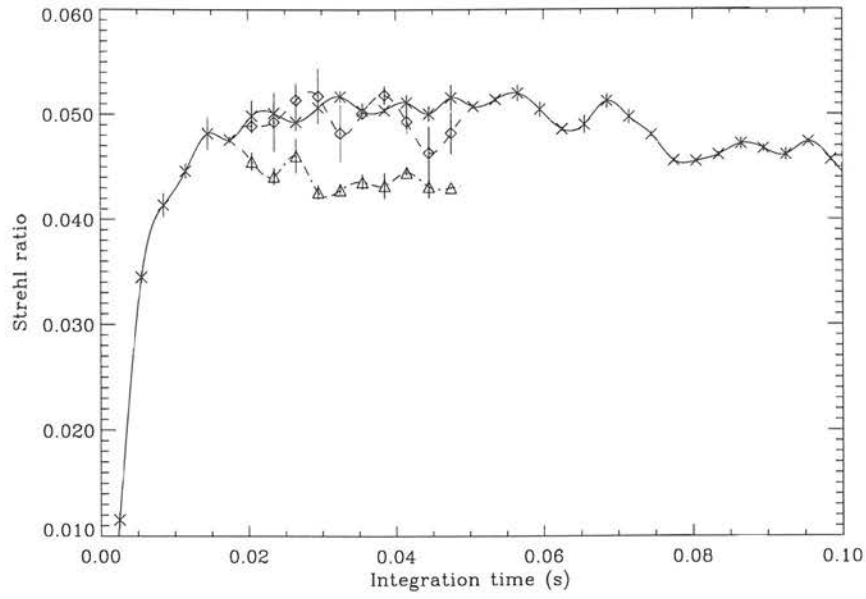


(a) Performance of system with atmosphere 1:
- ⟶ no prediction,
- -◇- neural networks and
- -△- linear predictors.

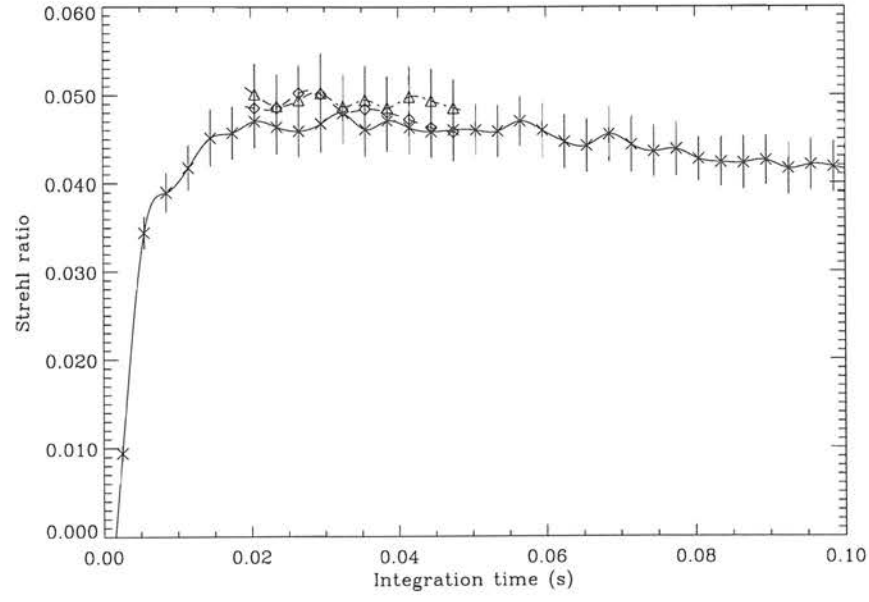


(b) As (a), but with atmosphere 2.

Figure 4.15: Investigation of 8×8 SH system performance with each phase screen with $r_0(0.55\mu m) = 15cm$.



(c) As (a), but with atmosphere 3.



(d) Performance averaged over the three atmospheres:
- ⟶ no prediction,
- -◇- neural networks and
- -△- linear predictors.

Figure 4.15: (*cont.*): Investigation of 8×8 SH system performance with each phase screen with $r_0(0.55\mu m) = 15cm$.

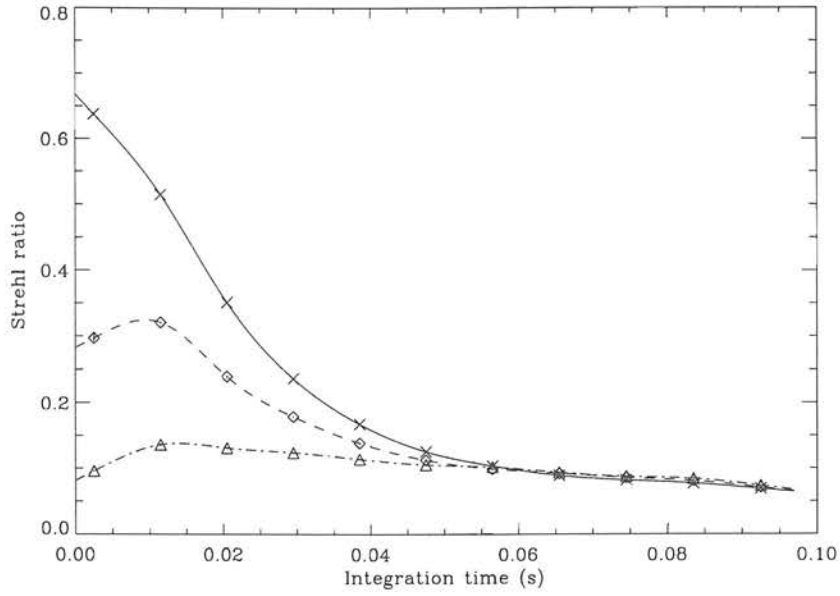
m_R	$t(ms)$	$\mathcal{S}_{AO}$	$\mathcal{S}_{NN}$	$\mathcal{S}_{LP}$
5.0	2.5	0.882 ± 0.001	–	–
10.0	2.5	0.872 ± 0.001	–	–
12.5	2.5	0.648 ± 0.005	0.662 ± 0.003	0.507 ± 0.010
13.5	10	0.367 ± 0.014	0.435 ± 0.010	0.323 ± 0.013
14.5	15	0.184 ± 0.014	0.226 ± 0.014	0.196 ± 0.012
15.5	35	0.117 ± 0.007	0.126 ± 0.009	0.131 ± 0.009
16.5	55	0.093 ± 0.006	0.096 ± 0.006	0.098 ± 0.006
17.5	75	0.084 ± 0.006	0.081 ± 0.005	0.082 ± 0.005

Table 4.2: Parameters and performance for each guide star magnitude.

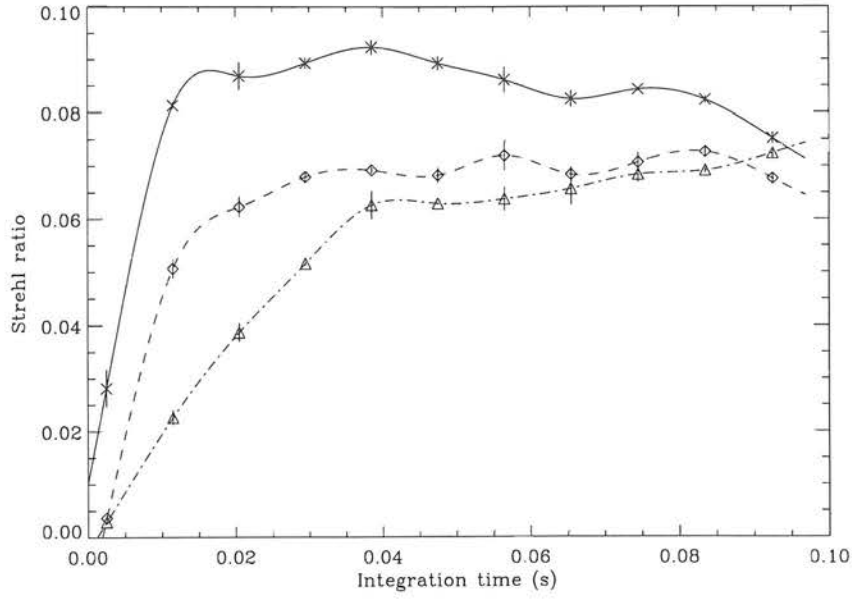
seeing of $r_0(0.55\mu m) = 30cm$ was selected. The predictor can be thought of as working in the background, observing what is going on and making suggestions which may or may not be accepted by the switch which has been introduced to the control loop.

A suitable integration time for the unaided AO system for each of the chosen magnitudes must be determined. The optimisation curves are for one atmospheric screen only to reduce runtime, and are shown in figure 4.16. The chosen magnitudes are tabulated in table 4.2 along with the selected WFS integration time.

For the performance graph, the total system time was $10s$ with performance data being collected for the last $9s$. The length of the learning set was kept at $0.5s$, with the servo warm–up being $0.25s$, again giving $0.25s$ of system time for convergence of the predictor to occur. Note that the number of modes which were predicted was still three.



(a) Performance of unaided AO system for various magnitudes:
- m_R= 12.5,
- m_R= 13.5 and
- m_R= 14.5.



(b) As (a), but with:
- m_R= 15.5,
- m_R= 16.5 and
- m_R= 17.5.

Figure 4.16: Optimisation curves for a realistic representation of an AO system.

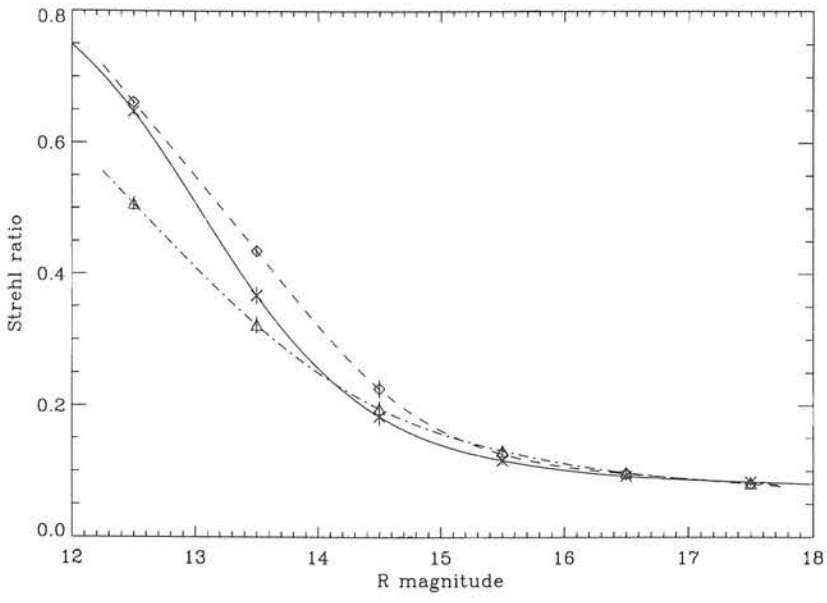
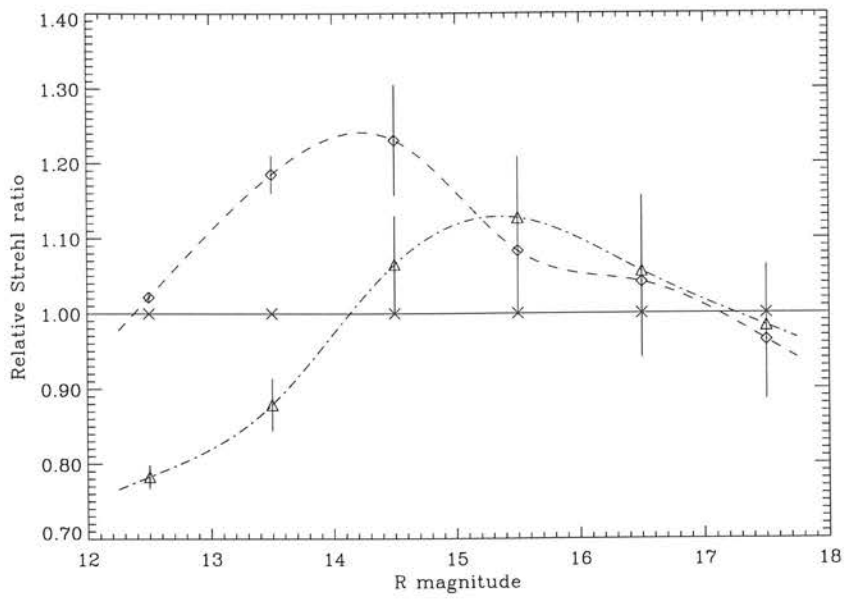


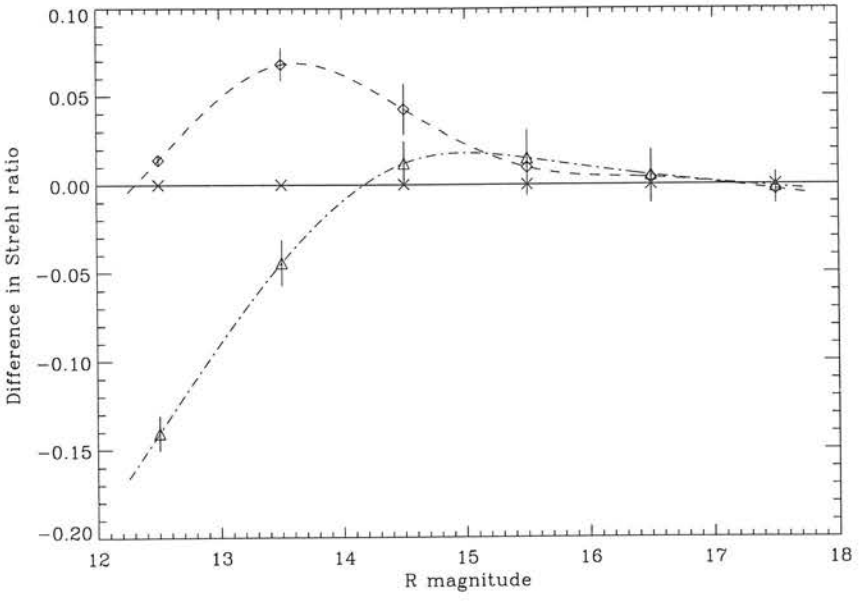
Figure 4.17: Comparison of system performance with magnitude:
-×- no prediction,
-◇- neural networks and
-△- linear predictors.

The results of this investigation are listed in table 4.2 and are graphed in figure 4.17. The time taken to obtain this data was about 3 weeks of real time or about 1 week of CPU time using a DEC AXP AlphaStation 500/400. The data are also displayed in other forms. Figure 4.18 (a) shows the data normalised by the unaided AO system curve and figure 4.18 (b) shows the data with the unaided AO system curve taken as the reference.

It is clear that the predictors of choice in this implementation scheme are the parallel neural networks. For the range $[12.5, 16.5]$ in guide star magnitude, the neural network predictors provide a benefit in system performance through reduction of the lag between measurement and correction. In this example, for magnitudes brighter than 12.5, the AO system is able to correct quickly enough



(a) As figure 4.17, but with curves divided by unaided AO case:
- ⟶ no prediction,
- -◇- neural networks and
- -△- linear predictors.



(b) As figure 4.17, but with unaided AO case subtracted from each curve:
- ⟶ no prediction,
- -◇- neural networks and
- -△- linear predictors.

Figure 4.18: Different presentations of data shown in figure 4.17.

so the lag effect is minimal and no performance benefit is expected through prediction. For a magnitude fainter than 16.5, noise begins to dominate, and the neural network curve joins the unaided AO system.

The parallel linear predictors do not fare so well in this implementation. At magnitudes brighter than 14.0, they cause the performance of the system to worsen. Between 14.0 and 16.5 a benefit is to be had through using the linear predictors, and for the range $[15.2, 16.5]$, linear predictors outperform neural networks to a small degree.

A possible explanation for this improvement at high magnitudes is that linear predictors are simpler beasts than neural networks. The neural networks may be fitting to the noisy fine structure while the linear predictors, which have less variables available for optimisation, can smooth the noise and therefore provide a better estimate of the underlying process.

In the case of the lower magnitudes, where the linear predictors are being beaten by the neural networks, the linear predictors may be performing too much smoothing of the data whereas the neural networks are still fitting to the fine structure. In this case, the fine structure is significant through good signal to noise, so the neural networks are tracking the process more accurately. The neural networks can therefore make better predictions which result in them outperforming the linear predictors.

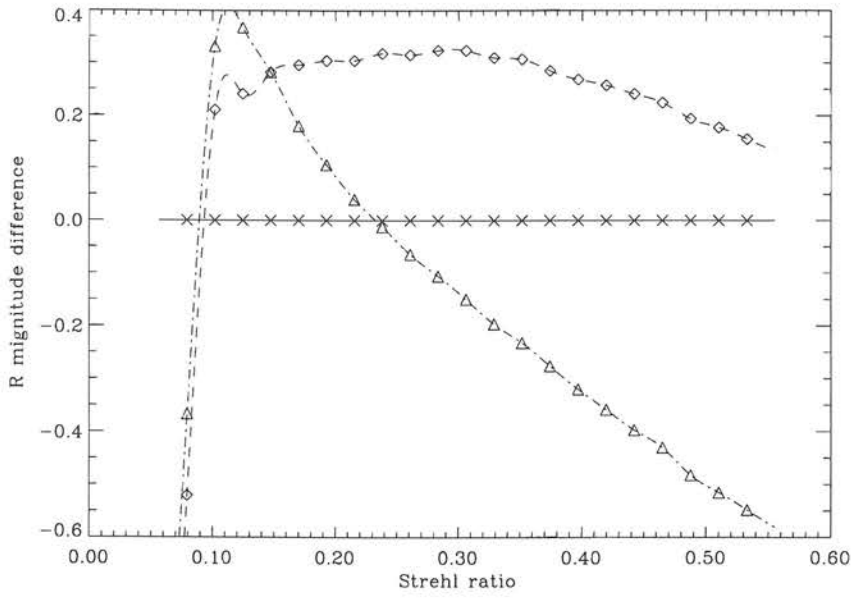


Figure 4.19: Increase in guide star magnitude for given Strehl:
-×- no prediction,
-◇- neural networks and
-△- linear predictors.

4.10.1 Fainter Stars for Same Strehl

An interesting result can be established if the data in figure 4.17 is plotted somewhat differently. For a range of Strehl ratios, the difference in the magnitude required to obtain that performance is calculated with the unaided AO system providing the baseline. This is graphed in figure 4.19 which shows that for a wide range of Strehl, the NN's allow a guide star of ~ 0.25 magnitudes fainter to be used to obtain the same performance of the AO system.

4.11 Available Increase in Sky Coverage

The final question to be asked is what benefit in sky coverage this increase in performance is likely to yield. Recall that in section 3.5.7 a result presented by Ellerbroek was highlighted, viz: "if one wishes to obtain a Strehl of 0.3 in the H near infrared band, there is about an 8% chance of finding a suitable reference star with which to guide an AO system. In the K band, this increases to $\sim$ 30%" (Ellerbroek et al., 1995).

The best improvement in Strehl obtained using the neural network case was about 0.06 at a magnitude of around 13.5. Using a neural network assisted AO system increases the performance of a system that was obtaining a Strehl of 0.25 to 0.3. In Ellerbroek's plot, the amount of sky with a suitable guide star at a Strehl of 0.25 is 10%. Introducing the predictive element to the control loop has therefore increased amount of sky viewable by adaptive optics. By a similar argument, the sky coverage at K increases to $\sim$ 40%.

By introducing the predictive element into the control loop to overcome the time lag inherent to an AO system, the area of sky available for use with an AO system may be increased by around 25%.

4.12 Comparison with Results from ONERA

As mentioned earlier, a group at ONERA has been working on a version of this problem (Dessenne et al., 1997; Dessenne et al., 1998). There are the following differences between our two implementations.

1. The algorithm developed at ONERA requires knowledge of the behaviour of the AO system. The method proposed here only requires that the mirror position be known accurately. Both methods allow the open loop case to be recovered from closed loop data.

2. ONERA use iterative optimisation of the coefficients a_i and b_i of a general linear controller of the form

$$\mathcal{C}(z) = \frac{\sum_{i=0}^{q-1} a_i z^{-i}}{1 + \sum_{i=0}^{p-1} b_i z^{-i}}. \tag{4.1}$$

   This formulation includes all possible linear predictors and is more general than the linear predictor discussed in section 3.9.1.

3. The atmospheric phase screens used in the respective simulations are very unlikely to be identical.

4. The ONERA implementation does not update the predictor in real time to reflect the most recent behaviour of the atmosphere.

5. The ONERA implementation does not contain a switch (section 3.10) — the controller/predictor is switched in at all times.

6. Although not implemented here, the quantities tested when setting the switch allow a real-time determination of the relative performance of the unaided and predictor aided AO system.

The differences in implementation are also reflected in the results the two techniques have obtained. ONERA report that their more general version of a linear controller/predictor is sufficient to make useful predictions, whereas the work here suggests that neural networks are a better choice than the more limited linear predictors implemented here.

It is possible that the neural networks implemented here could be equivalent to the more general linear controller/predictor used by ONERA. However, tests in which the switch in this implementation is forced to always select the neural network prediction have shown a slight decrease in performance compared with the unaided AO system.

4.13 Real System Implementation

The algorithm in chapter 3 has been shown to be beneficial in correcting for the time lag in an AO system. The very simple nature of the many parallel neural networks makes implementation of this technique relatively straightforward.

The retraining of the predictors after training set updates can take place on processors which are in parallel with the control loop processor. Predictor training

can therefore take place continuously, with the occasional interruption from the control loop processor when predictions are required. The only additions to the control loop programme are a relatively simple comparison stage, followed by prompting selected predictors for predictions.

There now follow some suggestions for possible improvements to the prediction scheme. Though possible to investigate numerically, they would be most efficiently investigated when implemented on a real system, as the time required for a significant sample of parameter space using the simulation is prohibitive.

4.13.1 Switching

As it stands, the switch in the control loop compares two predictions for the coefficient of each mode. If the predictor was performing better in the previous iteration, then it is switched in and used to define the next mirror shape. This is a hard switch. Depending on the properties of the atmosphere, it might be better to wait until two or more previous predictions have been shown to be better than the unaided system before switching in the predictor — a softer switch.

A softer switch, the precise characteristics of which could be based on histogram diagnostics (section 4.5), could remove the certain performance hit of a predictor switching in and being switched straight back out in the next iteration. However, this also reduces the benefit to be had from the clumps of consecutive prediction which are larger than 3. Careful selection of the control criteria can be used to

optimise the performance improvement to be obtained through prediction.

4.13.2 Many, Many Parallel Predictors.

The current implementation requires the switch to choose between two suggestions for each mode of the next mirror shape. One can envisage a selection of predictors, each more capable in a particular observing regime than the others (c.f. the LP's performing better than the NN's at faint magnitudes). The switch has the task of choosing between them.

4.13.3 Optimisation of Predictors

Similarly, predictor parameters such as the length of the training set and number of inputs per predictor can be continually tweaked to optimise the performance of the predictors.

4.13.4 Performance Monitoring

The knowledge of the difference in performance between a predictor and the unaided AO system's estimate for the mirror shape allows simultaneous estimation of the performance of the two cases. It is therefore possible to monitor the relative performance of unaided AO and predictor assisted AO.

4.14 Summary

In this chapter, the results of a number of runs of an AO simulation with the additional predictive element described in chapter 3 have been presented. It has been shown that it is necessary to use parallel modal predictors to obtain benefits with prediction. The degree of prediction can be significantly less than the degree of correction, provided that the prediction is performed on those mirror modes which have the greatest effect on image degradation. Using only 3 modes of prediction, an increase of $\sim$ 0.25 in guide star magnitude for similar Strehl has been found, enabling the area of the sky available for AO observation to be increased by $\sim$ 25%.

Chapter 5

Conclusions

The work presented in this thesis has been on two related infrared astronomical projects.

Chapter 2 detailed an observing programme to measure the first infrared trigonometric parallaxes. The accuracy to which the parallaxes were determined was $\sim$ 10 *mas*, which approaches that obtained with optical measurements. The introduction of larger detectors in conjunction with Tip/Tilt or low-order correction Adaptive Optics systems promises to allow even greater accuracy to be attained in the K-band.

An estimate of the possible benefits to parallax observations with such a system can be made. A factor of two increase in positional accuracy will come from smaller image sizes with a Tip/Tilt system. UKIRT can image at around 0.4″ FWHM with the Tip/Tilt system, as opposed to 0.7 $\sim$ 0.8″ sometimes obtained in observations for this work. Coupled with this, a second factor of two can be

obtained due to improved contrast for stars fainter than the background.

A third factor of two comes from using a 1024×1024 array. This will have 16 times the pixels compared to the 256×256 array in IRCAM3, but if a Tip/Tilt system is being used, the requirement to fully sample the point spread function means that only 4 times the area on the sky can be viewed. Thus, there will be roughly 4 times as many reference stars, with the expected increase in accuracy being $\sqrt{n_{stars}}$.

Overall, about an order of magnitude improvement in the accuracy of measured parallaxes should be possible with infrared techniques given current developments. Accuracy in the region of $\delta\pi = 1 \sim 2\ mas$ would then make infrared techniques directly comparable to optical observations.

Chapters 3 and 4 deal with adaptive optics, a system to compensate for atmospheric distortion when performing infrared astronomical observations. The number of bright stars in the sky limit the amount of sky which is observable with an AO system. A cause of the poorer performance of an AO system is the time lag between measurement and correction, during which time the atmosphere may have had time to evolve.

A novel technique to perform temporal prediction in a closed loop adaptive optics system was proposed, and numerical simulations performed to test the method. The method relies on the position of the deformable mirror in the system being

well known so that the behaviour of the atmosphere can be recovered. This is the case for the NAOMI system currently under development in the UK. Numerical simulations show that an improvement in performance can be obtained when using neural network predictors. This improved performance can be translated into an increase in the amount of sky observable with an AO system.

The next stage of the parallax work is to apply for service observations of both vB 10 and PC 0025+0447 using UFTI, the new imager for UKIRT. UFTI will fully sample the 0.4″ images which can be obtained using the Tip/Tilt system and these observations will allow the assertions regarding improvements in resolution to be tested.

The next step of the adaptive optics work is to implement the predictive scheme in the control loop of a system such as ELECTRA, a development system for NAOMI. ELECTRA and NAOMI share the same mirror. The efficacy of the algorithm could then be tested during the operation of a real system both in the laboratory and on a telescope. Another possible user of the algorithm could be the Tip/Tilt system in UKIRT.

In summary, this thesis has shown that it is possible to obtain infrared parallax estimates to good precision, and that with current innovations, infrared observations will become competitive with optical observations. One of the innovations currently in development is adaptive optics. A method has been proposed to improve the performance of AO systems using a temporal prediction algorithm.

Numerical simulations show that this algorithm could result in around 25% more sky being available to observers using AO in the infrared.

Bibliography

Aitken, G. J. M. and McGaughey, D.: 1995, in (Cullum, 1995), pp 89–94

Alloin, D. M. and Mariotti, J.-M. (eds.): 1993, *Adaptive Optics for Astronomy*, Vol. 423 of *NATO ASI Series*, Kluwer Academic Publishers

Angel, J. R. P., Wizinowich, P., Lloyd-Hart, M., and Sandler, D.: 1990, *Nature* **348**, 221

Babcock, H. W.: 1953, *P.A.S.P.* **65(386)**, 229

Beckenridge, J. B.: 1976, *J. Opt. Soc. Am.* **66(2)**, 143

Beckers, J. M.: 1993, *Ann. Rev. Astron. Astrophys.* **31**, 13

Boyd, R. W.: 1978, *J. Opt. Soc. Am.* **68(7)**, 877

Brase, J. M., An, J., Bissinger, H. D., Friedman, H. W., Gavel, D. T., Johnston, B., Max, C. E., Oliver, S. S., Presta, R., Rapp, D. A., Salmon, J. T., Waltjen, K. E., and Fisher, W.: 1994, in (Ealey and Merkle, 1994), pp 474–488

Buscher, D. F., Doel, A. P., Andrews, N., Dunlop, C., Morris, P. W., Myers, R. M., Sharples, R. M., Vick, A. J. A., Zadrozny, A., Haniff, C. A., and Wilson, R. W.: 1995, in (Cullum, 1995), pp 63–68

Close, L. M., Roddier, F., Roddier, C., Northcott, M., and Graves, J. E.: 1996, in *Adaptive Optics*, Vol. 13 of *Technical Digest Series*, pp 112–114, O.S.A.

Cullum, M. (ed.): 1995, *Adaptive Optics*, Vol. 54 of *Conference and Workshop Proceedings*, ESO

Dainty, J. C. and Scaddan, R. J.: 1975, *Mon. Not. R. Astr. Soc.* **170**, 519

Dessenne, C., Madec, P.-Y., and Rousset, G.: 1997, *Optics Letters* **22(20)**, 1535

Dessenne, C., Madec, P.-Y., and Rousset, G.: 1998, *Applied Optics* **37(21)**, 4623

Doel, P., Busher, D., Dunlop, C., Sharples, R., and Andrews, N.: 1995, in (Cullum, 1995), pp 69–74

Ealey, M. A. and Merkle, F. (eds.): 1994, *Adaptive Optics in Astronomy*, Vol. 2201 of *Proc. SPIE*

Ealey, M. A. and Washeba, J. F.: 1990, *Optical Engineering* **29(10)**, 1191

Ellerbroek, B. L., David W, T., and Racine, R.: 1995, in (Cullum, 1995), pp 357–362

ESA: 1997, *The Hipparcos and Tycho Catalogues*, European Space Agency

Freedman, H. W.: 1993, in (Alloin and Mariotti, 1993), pp 175–184

Fried, D. L.: 1965, *J. Opt. Soc. Am.* **55(11)**, 1427

Fugate, R. Q., Fried, D. L., Ameer, G. A., Boeke, B. R., Browne, S. L., Roberts, P. H., Ruane, R. E., Tyler, G. A., and Wopat, L. M.: 1991, *Nature* **353**, 144

Gallant, P. J. and Aitken, G. J. M.: 1996, in (Wizinowich et al., 1996b), pp 253–255

Gavel, D., Oliver, S., and Brase, J.: 1996, in (Wizinowich et al., 1996b), pp 21–23

Gendron, E. and Léna, P.: 1994, *Astron. Astrophys.* **291**, 337

Gilmore, G., Longmore, A., Myers, R., and Wilson, R.: 1995, in (Cullum, 1995), pp 9–15

Gliese, W. and Jahreiss, H.: 1991, *Preliminary Version of the Third Catalogue of Nearby Stars*, SIMBAD

Glindemann, A., Lane, R. G., and Dainty, J. C.: 1993, *Journal of Modern Optics* **40(12)**, 2381

Graham, J. R., Matthews, K., Greenstein, J. L., Neugebauer, G., Tinney, C. G., and Persson, S. E.: 1992, *Astron. J.* **104(5)**, 2016

Green, R. M.: 1985, *Spherical Astronomy*, Cambridge University Press

Greenwood, D. P. and Parenti, R. R.: 1993, in (Alloin and Mariotti, 1993), pp 185–204

Guinness: 1981, *Book of World Records* p. 295

Hambly, N. C., Miller, L., MacGillivray, H. T., Herd, J. T., and Cormack, W. A.: 1998, *Mon. Not. R. Astr. Soc.* **298**, 897

Harrington, R. S., Kallarakal, V. V., and Dahn, C. C.: 1983, *Astron. J* **88(7)**, 1038

Henry, T. J. and McCarthy, D. W.: 1993, *Astron. J.* **106(2)**, 773

Hertz, J., Krogh, A., and Palmer, R. G.: 1991, *Introduction to the Theory of Neural Computing*, Addison–Wesley

Hurburd, B. and Sandler, D.: 1990, *Optical Engineering* **29(10)**, 1186

Jones, H. R. A., Longmore, A. J., Jameson, R. F., and Mountain, C. M.: 1994, *Mon. Not. R. AStron. Soc.* **267**, 413

Jorgenson, M. B. and Aitken, G. J. M.: 1992, *Optics Letters* **17(7)**, 466

Kirkpatrick, J. D., Henry, T. J., and Liebert, J.: 1993, *Ap. J.* **406**, 701

Klückers, V. A., Wooder, N. J., Nicholls, T. W., Adcock, M. J., Munro, I., and

Dainty, J. C.: 1998, *Astron. Astrophys. Suppl. Ser.* **130**, 141

Kokorowski, S. A.: 1979, *J. Opt. Soc. Am.* **69(1)**, 181

Kolmogorov, A.: 1941, in D. G. Crowe (ed.), *Selected Papers on Adaptive Optics and Speckle Imaging*, Vol. 93 of *Milestone Series*, pp 3–7, SPIE

Lane, R. G., Glindemann, A., and Dainty, J. C.: 1992, *Waves in Random Media* **2**, 209

Leggett, S. K.: 1996, *Airmass correction in IRCAMDR*, Private communication

Leggett, S. K. and Hawkins, M. R. S.: 1988, *Mon. Not. R. Astr. Soc.* **234**, 1065

Linsky, J. L., Wood, B. E., Brown, A., Giampapa, M. S., and Ambruster, C.: 1995, *Ap. J.* **455**, 670

Marcus, S., Nelson, R., and Lynds, R.: 1979, in *Instrumentation in Astronomy III*, Vol. 172 of *Proc. SPIE*, pp 207–231

Martin, H. M.: 1987, *Pub. Astron. Soc. Pac.* **99**, 1360

McGaughey, D. R. and Aitken, G. J. M.: 1997, *J. Opt. Soc. Am. A* **14(8)**, 1967

McLean, I. S., Chuter, T. C., McCaughrean, M. J., and Rayner, J. T.: 1986, in D. L. Crawford (ed.), *Instrumentation in Astronomy IV*, Vol. 627 of *Proc. SPIE*, pp 430–437

Monet, D. G.: 1988, *Ann. Rev. Astron. Astrophys.* **26**, 413

Monet, D. G. and Dahn, C. C.: 1983, *Astron. J.* **88(10)**, 1489

Monet, D. G., Dahn, C. C., Vrba, F. J., Harris, H. C., Pier, J. R., Luginbuhl, C. B., and Ables, H. D.: 1992, *Astron. J.* **103(2)**, 638

Montera, D. A., Welsh, B. M., Roggemann, M. C., and Ruck, D. W.: 1996a, *Applied Optics* **35(21)**, 4238

Montera, D. A., Welsh, B. M., Roggemann, M. C., and Ruck, D. W.: 1996b, *Applied Optics* **35(29)**, 5747

Mould, J., Cohen, J., Oke, J. B., and Reid, N.: 1994, *A.J.* **107(6)**, 2222

Müller, B. and Reinhardt, J.: 1991, *Neural Networks: An Introduction*, Physics of Neural Networks, Springer–Verlag

Newton, I.: 1730, *Optiks*, , 4th edition

Nicholls, T. W., Boreman, G. D., and Dainty, J. C.: 1995, *Optics Letters* **20(24)**, 2460

Noll, R. J.: 1976, *J. Opt. Soc. Am.* **66(3)**, 207

Oliver, S. S., An, J., Avicola, K., Bissinger, H. D., Brase, J. M., Freidman, H. W., Gavel, D. T., Macintosh, B., Max, C. E., and Waltjen, K. E.: 1996, in (Wizinowich et al., 1996b), pp 34–36

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T.: 1988, *Numerical Recipies in C: The Art of Scientific Computing*, CUP

Racine, R., Salmon, D., Cowley, D., and Sovka, J.: 1991, *P.A.S.P.* **103**, 1020

Rigaut, F.: 1993, in (Alloin and Mariotti, 1993), pp 163–174

Rigaut, F., Ellerbroek, B. L., and Northcott, M. J.: 1997, *Applied Optics* **36(13)**, 2856

Rigaut, F., Salmon, D., Arsenault, R., McArthur, S., Thomas, J., Lai, O., Rouan, D., Gigan, P., Véran, J.-P., Crampton, D., Flatcher, M., Stilburn, J., Leckie, B., Roberts, S., Woof, R., Boyer, C., Jagourel, P., and Gaffard, J.-P.: 1996, in (Wizinowich et al., 1996b), pp 46–48

Roddier, F.: 1988, *Applied Optics* **27(7)**, 1223

Roddier, F.: 1993, in (Alloin and Mariotti, 1993), pp 89–112

Roddier, F., Northcott, M., and Graves, J. E.: 1991, *Pub. Astron. Soc. Pac.* **103**, 131

Roggemann, M. C., Welsh, B. M., Montera, D., and Rhoadarmer, T. A.: 1995, *Applied Optics* **34(20)**, 4037

Rousset, G., Beuzit, J.-L., Hubin, N., Gendron, E., Madec, P.-Y., Boyer, C., Gaffard, J.-P., Richard, J.-C., Vittot, M., P.Gigan, and Léna, P.: 1994, in (Ealey and Merkle, 1994)

Rowan-Robinson, M.: 1985, *The Cosmological Distance Ladder*, W. H. Freeman

RSI: 1994, *IDL User's Guide*, Research Systems, Inc.

Sarazin, M. and Roddier, F.: 1990, *Astron. Astrophys.* **227**, 294

Schneider, D. P., Greenstein, J. L., Schmidt, M., and Gunn, J. E.: 1991, *Astron. J.* **102(3)**, 1180

Schneider, D. P., Schmidt, M., Gunn, J. E., and Postman, M.: 1993, *P.A.S.P* **105**, 821

Steinhaus, E. and Lipson, S. G.: 1979, *J. Opt. Soc. Am.* **69(3)**, 468

Stetson, P. B.: 1987, *P.A.S.P.* **99**, 191

Tatarski, V. I.: 1967, *Wave propagation in a turbulent medium*, Dover

Thom, T.: 1987, *The Air Pilot's Manual*, Vol. 2, Airlife

Tinney, C. G.: 1993, *Astron. J.* **105(3)**, 1169

Tinney, C. G.: 1996, *Mon. Not. R. Astron. Soc.* **281**, 644

Tinney, C. G., Reid, I. N., Gizis, J., and Mould, J. R.: 1995, *Astron. J.* **110(6)**, 3014

Tyson, R. K.: 1991, *Principles of Adaptive Optics*, Academic Press

van Altena, W. F.: 1971, *A.J.* **76(9)**, 932

van Altena, W. F.: 1983, *Ann. Rev. Astron. Astrophys.* **21**, 131

van Biesbroek, G.: 1961, *A. J.* **66(9)**, 528

Vernin, J. and Muñoz–Tuñón, C.: 1994, *Astron. Astrophys.* **284**, 311

Wampler, S., Ellerbroek, B., and Gillies, K.: 1994, in (Ealey and Merkle, 1994), pp 239–243

Wild, W. J.: 1996, *Optics Letters* **21(18)**, 1433

Wilson, R. W. and Jenkins, C. R.: 1996, *Mon. Not. R. Astron. Soc.* **268**, 39

Wizinowich, P., Acton, D. S., Gleckler, A., Stomski, P., Avicola, K., Brase, J., Friedman, H., Gavel, D., and Max, C.: 1996a, in (Wizinowich et al., 1996b), pp 8–10

Wizinowich, P., Welsh, B., Ellerbroek, B., Bonaccini, D., Fender, J., Iye, M., Lelievre, G., Lloyd-Hart, M., Lukin, V. P., Myers, R., and Tyson, R. (eds.): 1996b, *Adaptive Optics*, Vol. 13 of *1996 Technical Digest Series*, OSA

Woolf, N. J.: 1982, *Ann. Rev. Astron. Astrophys.* **20**, 367

Wright, G. S.: 1998, *Inaccurate UKIRT encoders*, Private communication